



Microbial isoprenoid production

Carlsen, Simon

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Carlsen, S. (2012). *Microbial isoprenoid production*. Department of Systems Biology, Technical University of Denmark.

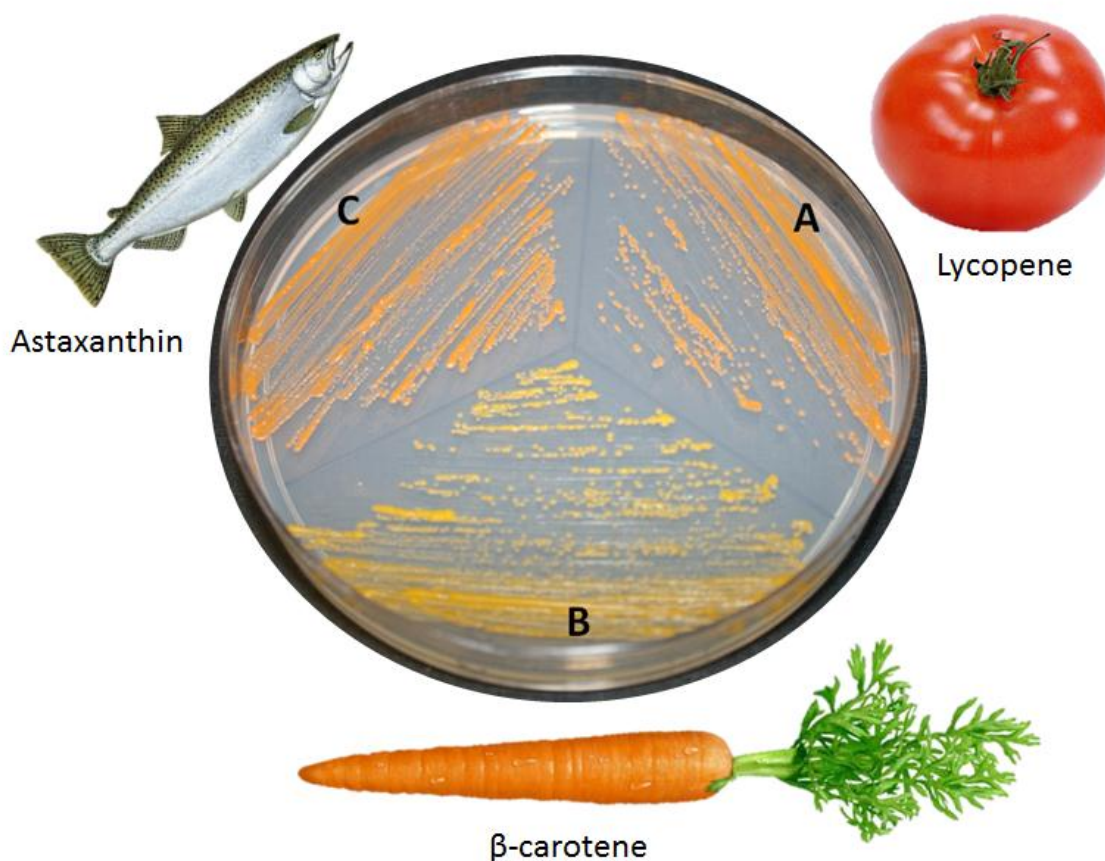
General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Microbial isoprenoid production



Simon Carlsen

Ph.D. Thesis

October 2012

Center for Microbial Biotechnology
Department of Systems Biology

Preface

The work presented in this thesis was carried out at Center for Microbial Biotechnology, Institute for Systems Biology, Technical University of Denmark and at Department of Chemical Engineering, Massachusetts Institute of Technology. The work was performed in the period from October 2008 to October 2012.

Dedicated posthumously to my grandfather, Victor Carlsen.

Date: 24th of September



Simon Carlsen

Acknowledgements

During the course of my Ph.D. studies, I have been fortunate to be able to collaborate and work with many very skilled and remarkable people who have helped, inspired, and supported me. I would therefore like to take this opportunity to extend my sincerest gratitude and appreciation to these persons.

First of all, I am very grateful for the financing provided by the Technical University of Denmark, which has enabled me to carry out my research in the distinguished scientific environment at the Center for Microbial Biotechnology (CMB), Department of Systems Biology.

I would like to express my deepest gratitude to my supervisors, Prof. Morten C. Kielland-Brandt, Associate Prof. Anna Elliason Lantz, and Dr. Michael Lyng Nielsen. Their help, guidance, encouragement, criticism, and support have been invaluable to me.

A large part of my Ph.D. was carried out in the laboratory of Prof. Gregory N. Stephanopoulos at the Department of Chemical Engineering, Massachusetts Institute of Technology (MIT). I would therefore also like to extend my most sincere appreciation to Prof. Stephanopoulos for letting me work in this very inspiring scientific environment. Furthermore, I would like to thank him for providing funding of my research, guidance, and inspiration.

From DTU, I would also like to thank Dr. Luca Riccardo Formenti, Dr. Mikael Rørdam Andersen, Dr. Jacob Blæsbjerg Nielsen, Dr. Kim Ottow, Dr. Linda Olkjær Lehmann, Dr. Kristian Fog Nielsen, and Mr. Bo Salomonsen for their advice, discussions of results, and suggestions. I would also like to extend special thanks to my office mates Mr. Stig Rattleff, Mr. Lars Poulsen, and Dr. Sujata Vijay Sohoni for creating a great work environment. Additionally I would like to thank the CMB technicians Ms. Jette Jepmond Mortensen, Ms. Tina Johansen, Ms. Elisabeth Krøger and Mr. Jesper Mogensen for their help and technical support. For help with the administrative part of my Ph.D. study, I would like to thank Ms. Birgitte Kolby H. Karsbøl, Ms. Dorthe Andersen, and Ms. Anne Nørrevang Johansen. Finally, I would like to thank all the employees at CMB; I am grateful for the incredible, social, fun and outgoing work place that you all have contributed to. They have made my time there amazing, especially the good times on Thursday afternoons.

While working at MIT, I had the privilege to work with some very skilled persons, who have helped me immensely; I would therefore like to extend my deepest appreciation to Assistant Prof. Marjan De Mey, Dr. Parayil Kumaran Ajikumar, Dr. Hang Zhou, Dr. Brian Pereira, Dr. Sarah M. Fendt, Dr. Haoran Zhang, Dr. Jie Zhang, Dr. Ryan Lim, Dr. Felix Lam, Dr. Jose Avalos, Mr. Adel Ghaderi, Dr. Mitchell Tai, Mr. Tom Wasylenko, Mr. Vikram G. Yadav, Dr. Kang Zhou. Of these persons, I would like to emphasize Mr. Adel Ghaderi and Dr. Brian Pereira, to whom I would like to extend a very special thanks, for besides being good colleagues also being really good friends. The rest of the Stephanopoulos lab also deserves my warmest

appreciation for contributing to creating an incredible work place, unparalleled by any other place in which I have ever worked.

While working at MIT, I was fortunate to be able to acquire plasmids and yeast strains from the comprehensive collection of Prof. Gerald Fink, Whitehead Institute, MIT. I would therefore like to thank Prof. Gerald Fink for granting me access to such an invaluable resource. This has greatly aided my research, and for this I am truly grateful.

Last but not least, I would like to express my deepest gratitude to my family and friends, who have supported me throughout the course of my Ph.D. studies. I would particularly like to thank my brothers, Mathias Carlsen and Andreas Carlsen and my parents Jane Boysen and Erik Carlsen, who have contributed invaluable help and support, especially during my stay abroad. Furthermore, having an active social life besides the studies and the academic community has been important to me. I am therefore very grateful to my friends: Jens, Jonas, Casper, Ulrik, Henrik, Asger, Casper, Karina. Thank you for all the good times!

Publications

Prior to starting my Ph.D. studies and during my Ph.D. studies I have co-authored two publications listed below that do not make part of this thesis. In addition, I am going to submit four articles based on chapters in this thesis. They have been listed below as well.

Publications:

Ajikumar PK, Tyo K, Carlsen S, Mucha O, Phon TH, Stephanopoulos G: Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharm.* 2008, 5(2):167–190.

Olsen LR, Hansen NB, Bonde MT, Genee HJ, Holm DK, Carlsen S, Hansen BG, Patil KR, Mortensen UH, Wernersson R: PHUSER (Primer Help for USER): a novel tool for USER fusion primer design. *Nucl. Acids Res.* 2011, 39: W61-W67.

Pending submissions:

Carlsen S, Nielsen ML, Lantz AE, Stephanopoulos GN, Kielland-Brandt MC: Heterologous microbial production of natural products: colorimetric surrogate metabolite screens.

Carlsen S, Parayil A, Formenti LR, Zhou K, Phon TH, Nielsen ML, Lantz AE, Kielland-Brandt MC, Stephanopoulos G: Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast.

Carlsen S, Nielsen ML, Kielland-Brandt MC, Stephanopoulos GN, Lantz AE: Image analysis and lycopene production as tools in semi-high-throughput screening for isopentenyl pyrophosphate overproducing colonies of *Saccharomyces cerevisiae*.

Carlsen S, Zhang H, De Mey M, Nielsen ML, Lantz AE, Kielland-Brandt MC, Stephanopoulos G: Engineering *Escherichia coli* for production of squalene: utilization of the *ERG9* gene from *Saccharomyces cerevisiae*.

Table of content

Preface.....	0
Acknowledgements	1
Publications	3
Publications.....	3
Pending submissions.....	3
Table of content.....	4
Summary.....	9
Dansk sammenfatning	12
Thesis outline.....	15
Aim of project.....	16
Chapter 1 – Heterologous microbial production of natural products: colorimetric surrogate molecule screens	18
1.1 Abstract.....	19
1.2 Introduction.....	20
1.3 Isoprenoids	22
1.4 Polyketides and flavonoids	27
1.5 Tools for improving throughput.....	29
1.6 Conclusions and future perspectives	31
1.7 Competing interests	32
1.8 References.....	33
Chapter 2 – Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in <i>Saccharomyces cerevisiae</i>: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast.....	40
2.1 Abstract.....	41
2.2 Introduction.....	42
2.3 Materials and Methods.....	49
2.3.1 Chemicals, enzymes and growth media.....	49
2.3.2 Plasmids and strains.....	49
2.3.3 Construction of expression cassettes	51
2.3.4 Construction of plasmids for integration of multiple expression cassettes	59
2.3.5 Construction of other plasmids.....	61
2.3.6 Construction of yeast strains	63

2.3.7	Test of expression cassette functionality using <i>yeGFP</i>	65
2.3.8	Liquid chromatography – Mass spectrometry analysis of MEP-pathway metabolites	65
2.3.9	Verification of pathway functionality using U-13C ₆ glucose.....	65
2.3.10	Verification of gene expression by reverse transcriptase polymerase chain reaction (RT-PCR)	65
2.3.11	Evolution of <i>S. cerevisiae</i> by serial sub-culturing.....	66
2.3.12	Microscopy.....	66
2.3.13	Functional complementation of <i>leu1</i> auxotrophy	66
2.4	Results and discussion	67
2.5	Acknowledgements.....	81
2.6	References.....	82
2.7	Supplementary material	88
2.7.1	Supplementary figure legends	88
2.7.2	Supplementary figures.....	89
Chapter 3 – Inverse metabolic engineering for increased production of isoprenoids in <i>Saccharomyces cerevisiae</i>: construction and screening of genomic DNA libraries		98
3.1	Abstract.....	99
3.2	Introduction.....	100
3.3	Materials and methods	107
3.3.1	Enzymes, chemicals and media.....	107
3.3.2	Plasmid and strain construction	107
3.3.3	Library construction	110
3.3.4	Library screening.....	111
3.3.5	Lycopene extraction, verification and quantification	111
3.4	Results and discussion	113
3.4.1	Strain construction and verification.....	113
3.4.2	Library construction	113
3.4.3	Library quality considerations.....	116
3.4.4	Library screening and strain testing.....	119
3.4.5	Alternative screens	121
3.5	Acknowledgements.....	123
3.6	References.....	124

Chapter 4 – Carotenoid biosynthesis in *Saccharomyces cerevisiae*: potential application of fluorescence-activated cell sorting for identification of isoprenoid hyper-producing mutants 130

4.1	Abstract	131
4.2	Introduction.....	132
4.3	Material and methods.....	137
4.3.1	Enzymes, chemicals and media.....	137
4.3.2	Plasmid and strain construction	137
4.3.3	Carotenoid extraction	141
4.3.4	Carotenoid analysis and quantification by HPLC	141
4.3.5	Fluorescence microscopy.....	142
4.3.6	Flow cytometry	142
4.4	Results and discussion	143
4.4.1	Establishing carotenoid biosynthesis in <i>S. cerevisiae</i>	143
4.4.2	Fluorescence microscopy.....	149
4.4.3	Flow cytometry of carotenoid producing <i>S. cerevisiae</i>	150
4.5	References.....	154

Chapter 5 – Image analysis and lycopene production as tools in semi-high-throughput screening for isopentenyl pyrophosphate overproducing colonies of *Saccharomyces cerevisiae* 158

5.1	Abstract	159
5.1.1	Background	159
5.1.2	Results.....	159
5.1.3	Conclusions	159
5.2	Background	160
5.3	Results.....	162
5.3.1	Construction of a yeast strain capable of controlled lycopene production.....	162
5.3.2	Image acquisition and analysis	164
5.3.3	Correlation of the yield of lycopene with red colony color intensity	168
5.3.4	cDNA library screening.....	172
5.4	Discussion	174
5.5	Conclusions.....	176
5.6	Methods.....	177
5.6.1	Enzymes, chemicals and media.....	177

5.6.2	Plasmids and strains.....	177
5.6.3	Image acquisition and analysis	180
5.6.4	Lycopene extraction and quantification	181
5.6.5	cDNA library construction and screening	181
5.7	List of abbreviations	183
5.8	Competing interests	183
5.9	Authors' contributions	183
5.10	Acknowledgements	183
5.11	References.....	184
6	Chapter 6 – Teaching old DNA new tricks: Engineering an <i>in vivo</i> transposon mutagenesis system for <i>Saccharomyces cerevisiae</i> based on the bacterial Tn5-transposase.....	186
6.1	Abstract.....	187
6.2	Introduction.....	188
6.3	Materials and methods	194
6.3.1	Enzymes, chemicals and media.....	194
6.3.2	Plasmid and strain construction	194
6.3.3	Component donor plasmids.....	197
6.3.4	Plasmids for verification of nuclear localization of the TNP transposase.....	197
6.3.5	Complete system	198
6.3.6	Strains	200
6.3.7	Nuclear localization.....	200
6.3.8	Plasposon system testing.....	201
6.4	Results and discussion	202
6.4.1	Nuclear localization.....	202
6.4.2	Plasposon system.....	204
6.4.3	Further experiments	209
6.5	References.....	210
	Chapter 7 – Engineering <i>Escherichia coli</i> for Production of Squalene: Utilization of the <i>ERG9</i> Gene from <i>Saccharomyces cerevisiae</i>	214
7.1	Abstract.....	215
7.2	Materials and methods	226
7.3	Acknowledgements.....	228

7.4	References.....	229
Chapter 8 – Conclusions and future perspectives.....		232
8.1	Chapter 2.....	234
8.2	Chapter 3, 4 and 5	236
8.3	Chapter 6.....	242
8.4	Chapter 7.....	243
8.5	Conclusion	245
8.6	References.....	246
Chapter 9 – Protocols		250
9.2	gDNA library construction protocol	251
9.2.1	Overview	251
9.2.2	Equipment, kits, enzymes, etc.:	252
9.2.3	Protocol.....	253
9.3	Yeast cDNA library construction protocol	256
9.3.1	Overview	256
9.3.2	Equipment, kits, enzymes, etc.:	256
9.3.3	Protocol.....	258

Summary

The main aim of this Ph.D. project is to engineer microorganisms to increase their production capabilities for the group of secondary metabolites known as isoprenoids and to develop and explore methods and approaches that may in the future be utilized to this end. The research was primarily focused on engineering *Saccharomyces cerevisiae*, but improved isoprenoid production in *Escherichia coli* was also addressed. We applied a two-pronged approach composed of rational and random engineering strategies. A summary of the aim of the research presented in each chapter combined with the primary results from this research can be found in the present section.

Chapter 1 is a minireview detailing the usage of colored metabolites as reporters for identification of microorganisms overproducing secondary metabolites. This chapter does not exclusively focus on isoprenoid reporter metabolites, but also reviews the application of colored polyketides and flavonoids. Former studies on the utilization of colored metabolite reporters are reviewed in conjunction with the potential applications of colored metabolites, which have recently been demonstrated to be heterologously produceable in microbes, for the setup of novel high-throughput screening methods. Because this review was written after finishing the result chapters of this thesis that are intended for publication, references to these results has been included. No further results generated as part of this Ph.D. project are presented within this chapter.

Chapter 2 encompasses our research into functional expression of the 2-C-methyl-D-erythritol (MEP)-pathway in *S. cerevisiae*. This has previously been attempted, but has probably failed due to problems of functionalizing heterologously expressed bacterial iron-sulfur cluster proteins within the yeast cytosol. We therefore attempted to graft the *E. coli* iron-sulfur cluster assembly machinery into *S. cerevisiae*. We were able to integrate the 16 genes, which we based on literature hypothesized would constitute this machinery, in a *S. cerevisiae* strain expressing the genes encoding the MEP-pathway. The results of HPLC-MS and labeling experiments showed that the pathway was indeed functional until the two last steps. These two steps are catalyzed by the iron-sulfur cluster proteins IspG and IspH. We finally demonstrated that *S. cerevisiae* is not completely unable to functionally express bacterial iron-sulfur cluster proteins in the cytosol by complementing a deficiency of the cytosolic iron-sulfur protein encoded by *LEU1* with the hetero-dimeric iron-sulfur cluster protein from *E. coli* encoded by *leuC* and *leuD*. It remains unknown why IspG and IspH were not functionally expressed in this organism.

Chapter 3 presents our research aiming to identify novel over-expression targets from gDNA libraries that can increase isoprenoid production in *S. cerevisiae*. We constructed a library based on *S. cerevisiae* gDNA integrated in a single-copy vector. The gDNA fragments were inserted downstream of the strong *TDH3* promoter to search for endogenous targets that upon over-expression could increase isoprenoid production. For screening large libraries an effective screen is imperative, so we engineered *S. cerevisiae* to produce the reporter metabolite lycopene. This

resulted in an orange-red phenotype enabling us to visually inspect colonies of library transformants, and hereby identify mutants with increased color intensity. We identified several mutants capable of increased lycopene production, but when the plasmids were rescued and re-transformed into *S. cerevisiae* only one of them was able to reconstitute the overproducing phenotype. When this plasmid was sequenced, it was found not to harbor an ORF. We further found that inspection through visual assessment of colony color was labor-intensive and difficult.

Chapter 4 presents our efforts towards establishing a high-throughput screening method based on the fluorescent carotenoids (a subgroup of the isoprenoids) β -carotene and astaxanthin, in conjunction with fluorescence activated cell sorting (FACS). We here demonstrate that the pathway resulting in lycopene formation (presented in Chapter 3) can be extended quite efficiently to yield β -carotene and astaxanthin producing *S. cerevisiae* strains; our engineered astaxanthin producing strain had an almost 9 fold higher astaxanthin yield compared to the best astaxanthin yield published for *S. cerevisiae*. When the carotenogenic *S. cerevisiae* strains were tested, it was found that the fluorescence was too weak to enable the application of FACS.

Chapter 5 describes the development of an image analysis algorithm for picking out the most intensively red colonies of lycopene producing *S. cerevisiae*. The algorithm is based on the analysis of digitally acquired agar plate pictures with the open source software ImageJ. We utilized the algorithm to pick 250 transformants from a transformation of lycopene producing *S. cerevisiae* with a *Saccharomyces kluyveri* cDNA library, and by comparing their ability to produce lycopene in liquid cultures to the same ability of 250 randomly picked transformants from the same population, we showed that the image analysis method was effective at picking out mutants with higher lycopene production.

Chapter 6 covers our effort towards the development of a novel transposon mutagenesis system for *in vivo* mutagenesis of *S. cerevisiae*. The system is based on the transposase from the bacterial Tn5 transposon. The system also encompasses a plasmid suicide mechanism based on the homing endonuclease I-SceI and the corresponding recognition site. The system is designed to insert a given DNA fragment randomly into the genome, and the preliminary results suggest that the system functions as intended, but further experiments are required.

Chapter 7 encompasses the construction and optimization of *E. coli* for production of squalene. Heterologous squalene production was mediated by the *S. cerevisiae* *ERG9* gene. To improve the production, we co-expressed the *iscA* gene from *E. coli*, encoding a FPP synthase. The plasmid harboring these two genes was tested in strains with different levels of overexpressions of four MEP-pathway genes. Based on the strain testing, we found that cell growth was affected by the heterologous production of squalene, which appears to limit the yield of squalene on biomass.

Chapter 8 discusses the results provided in Chapters 2 to 7 in combination with an overall conclusion intended to put this thesis into a wider perspective. Since most of the results have already been discussed and concluded upon in the individual chapters, much of the discussion

presented in Chapter 8 is of a more speculative character. In addition to the discussion and the overall conclusion, the chapter also describes a few experiments designed and performed after comparing the results from the result chapters. The additional results suggest that identification of novel genotypes causing increased isoprenoid production was hampered by an unintentional selection of respiratory deficient mutants.

Chapter 9 provides protocols for gDNA and cDNA library construction. The construction of high quality genetic libraries is not trivial and has been a big challenge, so to help others who have to perform library construction these protocols have been included. The protocols are more detailed than the descriptions found in the material and methods sections for Chapters 3 and 5 and should therefore be a better reference in the endeavor of genetic library construction.

Dansk sammenfatning

Hovedformålet med dette Ph.D. projekt er at konstruere mikroorganismer med øget produktionskapacitet for gruppen af sekundære metabolitter kendt som isoprenoider, samt at udvikle og undersøge metoder og fremgangsmåder, som har dette endemål. Arbejdet har primært været fokuseret på *Saccharomyces cerevisiae*, men det har også omfattet forbedret isoprenoid-produktion i *Escherichia coli*. Mine medforfattere og jeg har anvendt en fremgangsmåde bestående af både rationelle og tilfældige strategier. I nærværende afsnit har jeg sammenfattet målene for forskningen præsenteret i hvert kapitel samt resultaterne deraf.

Kapitel 1 er et minireview, som beskriver anvendelsen af farvede metabolitter, som rapportører til identifikation af mikroorganismer, som overproducerer sekundære metabolitter. Dette kapitel omfatter ikke kun isoprenoider, men også anvendelsen af farvede polyketider og flavonoider. Tidligere studier, der har gjort brug af farvede metabolitter, bliver her gennemgået sammen med potentiel anvendelse af andre farvede metabolitter, som nyligt er fundet mulige at producere heterologt i mikroorganismer, til at konstruere nye høj-effektive screenings-metoder. Eftersom kapitel 1 er blevet udarbejdet efter færdiggørelsen af de øvrige af de kapitler i denne afhandling, som er skrevet med henblik på publication, vil referencer til disse være inkluderet. Ingen andre resultater genereret i dette Ph.D. projekt præsenteres i dette kapitel.

Kapitel 2 inkluderer vores forskning i funktionel ekspression af 2-C-methyl-D-erythritol (MEP)-biosyntesevejen i *S. cerevisiae*. Dette er tidligere blevet forsøgt, men ser ud til at være mislykket på grund af problemer med funktionel ekspression af heterologe bakterielle FeS-cluster-proteiner i gærs cytosol. Vi forsøgte derfor at overføre maskineriet for konstruktion af FeS-clusters fra *E. coli* til *S. cerevisiae*. I en *S. cerevisiae* stamme, der i forvejen var konstrueret til at udtrykke de gener, som koder for MEP-biosyntesevejen, integrerede vi således 16 gener, som vi med baggrund i litteraturen antog, ville udgøre dette maskineri. Resultaterne af HPLC-MS og mærkningsforsøg viste, at biosyntesevejen netop var funktionel, indtil de reaktioner, der katalyseres af FeS-cluster-proteinerne, IspG og IspH, men ikke fungerede. Endeligt viste vi, at *S. cerevisiae* ikke er helt ude af stand til funktionelt at udtrykke bakterielle FeS-cluster-proteiner i sit cytoplasma ved at komplementere en mangel på det cytosoliske FeS-cluster-protein kodet af *LEU1* med det heterodimeriske protein fra *E. coli* kodet af *leuC* samt *leuD*. Det er stadig uvist, hvorfor IspG og IspH ikke kan udtrykkes funktionelt.

Kapitel 3 præsenterer vores forsøg på identifikation af *S. cerevisiae* gener, som ved overudtrykkelse kan øge produktionen af isoprenoider i *S. cerevisiae*. Vi konstruerede et bibliotek baseret på *S. cerevisiae* gDNA indsat efter den stærke *TDH3* promotor i et lav-kopi plasmid. For at kunne screene et så stort bibliotek effektivt konstruerede vi en *S. cerevisiae* stamme, som syntetiserer reporter-metabolitten lycopon, som resulterer i at cellerne får en orange-rød farve. Vi kunne herved visuelt indentificere adskillige kolonier af gærceller, som producerede en øget mængde af isoprenoider, men når vi genudvandt plasmiderne fra disse transformanter og benyttede dem til at re-transformere *S. cerevisiae*, kunne kun et enkelt af disse

genskabe overproduktions-fænotypen. Da dette plasmid blev sekventeret, viste det sig, at insertionen ikke indeholdt en ORF. Vi fandt desuden, at den visuelle inspektion af kolonifarve var vanskelig og arbejdskrævende.

Kapitel 4 præsenterer vores arbejde på at etablere en effektiv screenings metode baseret på de fluorescerende carotenoider (en undergruppe af isoprenoiderne) β -caroten og astaxanthin i samspil med "fluorescence-activated cell sorting" (FACS). Vi viser her, at biosyntesevejen, som fører til dannelse af lycopen (jf. kapitel 3), kan forlænges så *S. cerevisiae* kan producere β -caroten og astaxanthin med god effektivitet. Vores astaxanthin-producerende stamme gav et 9 gange højere astaxanthin-udbytte sammenlignet med det bedste udbytte offentliggjort for *S. cerevisiae*. Da de carotenoid-producerende stammer blev testet med FACS, viste det sig, at fluorescens-niveauet var for lavt til, at denne metode kan anvendes som screeningsmetode.

Kapitel 5 beskriver udviklingen af en algoritme for billedanalyse til udvælgelse af de mest intenst røde kolonier af *S. cerevisiae*. Algoritmen er baseret på analyse af digitale billeder af kolonier på agar plader med open-source programmet ImageJ. Vi anvendte algoritmen til at udvælge 250 transformanter fra en konstrueret lycopene-producerende *S. cerevisiae* stamme med et *Saccharomyces kluyveri* cDNA bibliotek. Vi sammenlignede disses evne til at producere lycopen i flydende medium med same egenskab ved 250 kolonier, som blev udvalgt tilfældigt fra den samme population. Vi kunne herved vise, at billedanalysen var i stand til at udvælge transformanter med forøget produktion af lycopen.

Kapitel 6 beskriver vores indsats henimod udviklingen af et nyt system for *in vivo* transposon-mutagenese af *S. cerevisiae*. Systemet er baseret på transposasen fra det bakterielle Tn5 transposon. Det indeholder endvidere en plasmid-"suicide-mechanism" baseret på I-SceI homing endonukleasen og den hertil hørende genkendelses-sekvens. Systemet er designet til at integrere en given DNA-sekvens et tilfældigt sted i genomet, og de foreløbige resultater antyder, at systemet virker efter hensigten, men flere eksperimenter er påkrævet.

Kapitel 7 omfatter konstruktion og optimering af en *E. coli* stamme til produktion af squalen. Heterolog squalen produktion er opnået ved hjælp af *ERG9* genet fra *S. cerevisiae*. Vi indførte desuden *iscA* genet fra *E. coli*, som koder for en FPP synthase, for at forbedre produktionen yderligere. Plasmidet, som huser disse to gener, blev testet i stammer med forskellige niveauer af over-udtrykkelse af fire gener i syntesevej til IPP. På denne måde fandt vi, at cellevæksten blev påvirket negativt af den heterolog produktion af squalen, hvorved udbyttet af squalen på biomasse ser ud til at blive begrænset.

Kapitel 8 diskuterer resultaterne fremlagt i kapitel 2 til 7, kombineret med en overordnet konklusion, som har til hensigt at sætte hele afhandlingen ind i et større perspektiv. Eftersom de fleste af resultaterne allerede har været diskuteret og konkluderet på i de individuelle kapitler, har det været naturligt at udvide diskussionen præsenteret i kapitel 8 til at være af en mere spekulativ karakter. Udover denne diskussion og den overordnede konklusion indeholder kapitlet

også nogle få resultater fra eksperimenter, som blev designet og udført, efter at resultaterne fra de andre kapitler blev sammenlignet. Disse resultater peger på, at identifikationen af nye genotyper, som øger produktionen af isoprenoider, blev vanskeliggjort af en utilsigtet udvælgelse af respiratoriske defekte mutanter.

Kapitel 9 giver protokoller for konstruktion af gDNA og cDNA bibliotekerne. Konstruktion af genetiske biblioteker af høj kvalitet er ikke trivielt og har været en stor udfordring. Så for at hjælpe andre, som skal konstruerer biblioteker, er disse protokoller blevet inkluderet. Protokollerne er mere detaljerede end de beskrivelser, som findes i material- og metode-afsnittene i kapitlerne 3 og 5, og de udgør hermed en bedre reference til konstruktion af sådanne genetiske biblioteker.

Thesis outline

The main topic of this thesis is heterologous microbial production of natural products with focus on isoprenoids. This thesis is organized in 9 chapters. Since I am planning to publish 4 of the 9 chapters in scientific journals, I found it expedient to write them all in article format. Because the chapters are intended to be published in different journals, the formatting varies in some chapters, e.g. the method of citation. The article format also implies that each chapter contains its own introduction, which, due to extensive overlap in topic between chapters, will contain some redundancies. Due to this structure, and even more due to the fact that metabolic engineering and isoprenoids have been extensively reviewed in recent years, both in the scientific literature and by my predecessors working within these areas of research at CMB, I have chosen not to include a regular introduction in this thesis. I have instead chosen to write a mini-review (Chapter 1) regarding colorimetric surrogate metabolite screening, which will serve as introduction to the part of the thesis encompassing random approaches. After this initial chapter follows the first of two chapters where rational engineering strategies have been applied. The introduction to this chapter (Chapter 2) is thorough and detailed and gives a good introduction to isoprenoid biosynthesis. These two parts of the thesis will, in addition to the introduction to each chapter, therefore serve as the prerequisite knowledge offered to the reader; if more extensive background knowledge is required and to avoid the creation of further redundancies, the reader is referred to reviews and other literature cited in the thesis.

Chapter 2 describes the efforts towards expressing the MEP-pathway and the expression of *Escherichia coli* cytosolic iron-sulfur cluster proteins in *Saccharomyces cerevisiae*, and it provides evidence that *S. cerevisiae* can functionally express at least one bacterial iron-sulfur cluster proteins within its cytosol. Chapters 3-6 encompass the efforts towards identification of novel genetic targets for improving isoprenoid biosynthesis in *S. cerevisiae* through random approaches. In Chapter 7, the rational engineering efforts for the construction of an *E. coli* strain for heterologous production of squalene is described. These six chapters comprise the experimental part of this thesis.

To conclude the thesis, a chapter with future perspectives, suggestions for additional work and conclusions is included. This chapter is subdivided into sections based on the prior chapters in which experimental work is presented. Since the results of the individual chapters have been discussed herein, this section will be of more speculative character. The final chapter of the thesis contains protocols for gDNA and cDNA library constructions, since this has been one of the major challenges that I have faced during my laboratory work; by including these protocols I hope to enable others to easier construct these two types of libraries.

Aim of project

Due to declining drug discovery rates from organic synthetic libraries, the pharmaceutical companies are again turning their attention towards secondary metabolites. Secondary metabolites have evolved to interact with other biomolecules, and the probability of identifying novel drugs from secondary metabolite libraries is therefore estimated to be much higher compared to the screening of synthetic libraries. However, secondary metabolites are in most cases only scarcely available in the native production organisms and need to be produced in an alternative way to large-scale extraction. The most common way of producing these compounds is by organic synthesis. However, organic synthesis can have serious disadvantages for production of secondary metabolites, such as low yields due to the complex structures, which can make this way of production economically unfeasible. An attractive alternative for producing secondary metabolites is to engineer microorganisms for heterologous production of the desired compound by transfer of biosynthetic genes. It is generally straightforward to scale up microbial production to meet actual demands, and it is also an environmentally benign production method compared to organic synthesis.

Among isoprenoids only high value chemical entities such as pharmaceuticals can currently be produced cost-effectively in microorganisms, but concurrently with increasing oil prices and the development of more effective microbial production methods the microbial production of more general chemicals will be feasible due to improved cost-competitiveness. By improving microorganisms through engineering, we can facilitate the shift towards a more bio-based economy hereby liberating the production of chemical entities from petrochemical raw materials by substitution with renewables resources.

Isoprenoids, also known as terpenoids, constitute the largest group of secondary metabolites identified to date, encompassing more than 55,000 different compounds including several blockbuster drugs such as paclitaxel and artemisinin. These compounds are all biosynthesized from the same precursor, called isopentenyl pyrophosphate (IPP), which is stepwise polymerized and diversified, giving rise to the enormous chemical and structural diversity. Thus it would be attractive to engineer a microorganism to produce high amounts of IPP and other immediate prenyl precursors such as geranyl pyrophosphate, farnesyl pyrophosphate and geranylgeranyl pyrophosphate; this organism could serve as cell factory for large-scale microbial production of isoprenoids.

Microorganisms have evolved to adapt to different niches in nature and are therefore not generally optimized for the purpose of production of a specific secondary metabolite. Optimization of microorganisms can be a daunting task due to the complexity of metabolism. In order to optimize microorganisms, metabolic engineering was conceived almost 20 years ago as a systematic methodology designed to deal with the inherent problems arising from this complexity by applying iterative steps of modeling, characterization and rationally designed perturbations. Numerous examples of its application in optimizing microorganisms for secondary

metabolite production can be found in the literature. One of the most often applied microorganisms for isoprenoid production is *Saccharomyces cerevisiae*. Several metabolic engineering approaches for production of isoprenoids in *S. cerevisiae* have been published, but most of them are based on few types of perturbations. The identification of novel genetic targets is a prerequisite for the continued improvement of microbial isoprenoid production. The main aim of this project is therefore to optimize *S. cerevisiae* for isoprenoid production through identification of novel targets and pathway engineering. Furthermore, it is the aim of this thesis to develop and devise methodologies that will further the main aim.

Chapter 1 – Heterologous microbial production of natural products: colorimetric surrogate metabolite screens

Simon Carlsen^{1,2}, Michael Lyng Nielsen^{2,3}, Anna Eliasson Lantz^{2,§}, Gregory N. Stephanopoulos¹, Morten C. Kielland-Brandt²

[§] Corresponding author.

¹ Department of Chemical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA.

² Department of Systems Biology, Technical University of Denmark (DTU), DK-2800 Kongens Lyngby, Denmark.

³ Present address: Novozymes A/S, Krogshøjvej 36, DK-2880 Bagsvaerd, Denmark.

1.1 Abstract

Heterologous microbial production of natural products has been one of the focal points of biotechnology for the last three decades and has therefore been extensively reviewed and the introductory part of this review offers a short overview of these, before focusing on the application of colored metabolites in surrogate metabolite screens. We find that this niche within screening methodologies deserves more attention than it has had in previous reviews. The review is divided into sections based on the major groups of metabolites for which the individual surrogate metabolites have been applied. Furthermore, a section focusing on approaches for automation applicable for improving the throughput of these screening methods is included. Finally, criteria for a good surrogate metabolite is addressed, together with future perspectives of applying these screens in the endeavor of creating microbial cell factories for production of natural products.

1.2 Introduction

Production of chemical compounds has to a large extent been based on fossil oil, but concurrently with rising oil prices and an increasing demand for specialty chemicals, heterologous microbial production is becoming an attractive alternative. The industrial and pharmaceutical biotechnology markets are therefore predicted to increase significantly over the next decade [Meyer and Turner 2009]. Several small molecule products and pharmaceuticals are already produced by microorganisms today, and microbial production generally offers several advantages, e.g. easy scale-up, chiral pure compounds can be obtained, environmentally benign production, and sometimes an opportunity to utilize inexpensive carbon sources such as waste products from other industries. Heterologous microbial production is also gaining further footing due to the renaissance of natural products as screening resource for pharmaceuticals [Dove 2003; Harvey 2007]. We have during the last couple of decades witnessed the initial steps towards a bio-based economy in which industrial biotechnology has a central role, in particular in the utilization of microbes to produce a wide range of chemical entities. This trend will very likely continue in the years to come [Octave and Thomas 2009].

Most microorganisms are not naturally optimized for production purposes, so after production of a desired molecule has been established, the given microorganism generally requires engineering to make the production as cost-effective as possible. For a general review on microbial strain improvement and the optimization of fermentative production see Parekh et al. (2000). Cultivation of some microorganisms in large tanks can be troublesome and many microorganisms have not been developed for genetic engineering, hampering their utilization [Chiang 2004]. However, by identifying the genetic basis for the production of the desired molecule within its natural host the production capability may be transferred to a heterologous microorganism for which genetic engineering, cultivation techniques and scale-up is well established. Some source organisms, such as many plants, suffer from other drawbacks, e.g. scarce availability of the desired molecule within the native producer and seasonal variations affecting production yields. These issues may also be solved by heterologous production in microorganisms. The established production microorganisms can then subsequently be optimized to create a microbial production platform.

Several methodologies have been developed to improve the production capabilities of microorganisms. These methodologies can be divided into random/combinatorial and rational approaches. Available knowledge on the system chosen for engineering is a prerequisite for the application of any rational approach. Knowledge on microbial systems is gradually increasing, but random approaches have nonetheless been the prevalent method used for improving microorganisms. For a review on classical strain improvement see Rowlands (1984). The advent of modern genetics and microbiological techniques paved the way for the invention of directed evolution in the 1960's [Mills et al. 1967]. Directed evolution can be applied to single enzymes, pathways or whole cells and can be regarded as a fast and highly directional version of

Darwinian evolution. Directed evolution consists of two parts; the creation of diversity and the screening or selection of the diversified population. Several methods for creating diversity within microbial population have been developed, and for comprehensive reviews on this topic the reader is referred to Zhao et al. (2002) and Chatterjee and Yuan (2006).

The success of any random approach is dependent on the efficacy of the method applied for identifying the microbial cells with beneficial mutations. Identification of mutants from libraries can be performed by screening the population for a certain phenotype or applying a selection pressure that eliminates microbes not harboring the desired traits. Selecting for a certain phenotype is often a more powerful method than screening since the microbes do not have to be assessed individually. However, most phenotypes, e.g. many over-production phenotypes, are not easily amenable to selection for viability or growth and must therefore be identified through screening. The phenotypes that are most straightforward to screen are discrete phenotypes, i.e. those that can be screened in an “either-or” fashion. These screens will however only permit the investigator to identify all the mutants with a certain trait and not enable the identification of the best mutant within this population. On the other hand, screening for continuous phenotypes can be difficult, since a genotypic change might only manifest itself as a subtle change in phenotype. For general reviews on screening methods, the reader is referred to Olsen et al. (2000) and Reymond and Babiak (2007).

Not all phenotypes can be readily screened for [Tyo et al. 2007], so in order to enable the construction of strains with high production capabilities, surrogate compounds synthesized from the same precursor metabolite as the desired compound can be applied to establish microbial production platforms with high flux towards the precursor molecule, increase the performance of a single enzyme, or improve the flux through an entire pathway. Screens for surrogate compounds can be based on various physical traits, e.g. fluorescence. This review presents the application of surrogate molecules for colorimetric screening of microbes with improved capabilities, with focus on heterologous production of natural products.

1.3 Isoprenoids

More than 55,000 isoprenoids have been identified, making isoprenoids the largest group of natural products. This group of ubiquitous molecules holds a great potential for the isolation of compounds with novel activities useful to mankind, as witnessed by, e.g. paclitaxel (Taxol®) and artemisinin, two of the most pronounced blockbuster drugs identified within recent time, for treatment of cancer and malaria, respectively. All natural isoprenoids are produced from the universal precursor isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP). DMAPP is polymerized with different numbers of IPP molecules by length-specific prenyl transferases to form prenyl pyrophosphate molecules of various lengths, which undergo further modifications to form the wealth of isoprenoid molecules from the smallest, e.g. isoprene (from a single DMAPP molecule), to the largest, e.g. poly-isoprenoids, such as natural rubber, of which each molecule is derived from more than 1000 IPP molecules. For a review on isoprenoids, their biosynthesis, and heterologous production, the reader is referred to Ajikumar et al. (2008).

The C-40 group of isoprenoid molecules known as carotenoids has been extensively used as pigments in colorimetric screens to improve heterologous isoprenoid production. Within this group of secondary metabolites, lycopene is the pigment with the shortest biosynthetic pathway from IPP and has been most commonly applied in screening methods. Ohnuma et al. (1994) utilized heterologous production of lycopene in *Escherichia coli* to set up a color complementation screen which they coined red-white screening. Lycopene is synthesized from the prenyl pyrophosphate known as geranylgeranyl pyrophosphate (GGPP), which is not naturally synthesized by *E. coli*. By omitting the GGPP synthase they integrated a heterologous pathway with a “gap”, allowing them to screen libraries in a discrete manner for novel enzymes with GGPP synthase activity. By applying this screening system and picking the colonies exhibiting the red lycopene production phenotype, they were able to isolate a novel GGPP synthase gene from a genomic DNA (gDNA) library from *Sulfolobus acidocaldarius*. Ohnuma et al. [1996] applied the same screening method in the creation of a novel GGPP synthase from the *Bacillus Stearothermophilus* farnesyl pyrophosphate (FPP) synthase. Following NaNO₂ random mutagenesis they picked red colonies harboring mutated versions of the gene normally encoding FPP synthase activity now encoding GGPP synthase activity. By assessing the various mutations resulting in amino acid substitutions they were able to gain insight in moieties involved in determination of prenyl pyrophosphate chain length. A similar study was carried out by Lee et al. (2005), but instead of mutating a heterologous GGPP synthase they targeted the *E. coli* FPP synthase encoded by *ispA*, which they mutated by error-prone PCR. Furthermore, instead of using lycopene they utilized a mutated phytoene synthase which catalyzes the formation of two additional double bonds to yield tetrahydrolycopene which supposedly is more readily observed on a white filter background [Lee et al. 2005]. With the aim of improving heterologous production of the isoprenoid levopimaradiene in *E. coli*, Leonard et al. (2010) also targeted the activity of a GGPP synthase. They opted to use the *Taxus canadensis* GGPP synthase which was

modified by error-prone PCR, and its activity was screened by heterologous production of lycopene. As opposed to the previous studies where the screening was carried out in a discrete manner, Leonard et al. (2010) were interested in a GGPP synthase with increased flux towards GGPP and therefore had to apply a continuous screen, screening for colonies with more intense red coloration. With visual inspection, this screen is prone to mistakes, and further testing of the selected mutants did indeed reveal false positives. Nevertheless, the authors were able to isolate a GGPP synthase which when co-expressed with the native levopimaradiene synthase resulted in an 18 fold increase in levopimaradiene production.

The carotenoid pathway can be readily extended to produce other carotenoids of commercial interest such as astaxanthin. Astaxanthin is in *Brevundimonas* produced from lycopene through β -carotene, which undergoes a series of oxidations catalyzed by 4,4'- β -ionone ring ketolase (CrtW) and 3,3'- β -ionone ring hydroxylase (CrtZ). In these oxidations, CrtW has been found to be the limiting step. To remedy this, Tao et al. (2006) sought to identify novel versions of CrtW and to modify these to improve the production of astaxanthin. They utilized the color change in phenotype brought upon *E. coli* heterologously expressing either β -carotene or hydroxylated versions hereof (yellow) compared to astaxanthin or the keto-derivatives of β -carotene (orange/reddish). The *Sphingomonas* sp. DC 18 isolated from freshwater (98 % similar to *Sphingomonas melonis* by 16S rDNA comparison) served as donor for the construction of a gDNA library. The library was screened in an *E. coli* strain heterologously producing β -carotene by selecting the colonies exhibiting an orange phenotype compared to the original yellow one. From the isolated colonies it was possible to obtain a gene encoding CrtW. When this new gene was co-expressed with *crtZ* in a β -carotene producing *E. coli* strain, the resulting strain produced 12 % astaxanthin, 88 % intermediates between β -carotene and astaxanthin. By mutating the *crtW* gene using error-prone PCR and picking colonies with increased red colony coloration it was possible to obtain mutants producing an increased proportion of astaxanthin. The best mutant obtained produced 83 % astaxanthin. Besides improvement of enzyme functionality and carotenoid production, colorimetric screens have also been utilized for isolation of genes producing non-native carotenoids such as the aforementioned tetrahydrolycopene and in combinatorial studies taking advantage of the promiscuous nature of carotenoid biosynthetic enzymes to probe carotenoid molecular space. For a comprehensive review of this work see Umeno et al. (2005).

The so far mentioned examples of screens have primarily been concerned with improving pathways downstream of IPP; however increasing flux towards IPP is of outmost importance in construction of microbial cell factories for production of isoprenoids. IPP is biosynthesized through two alternative pathways, namely the mevalonate pathway and the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. Random and combinatorial approaches where colorimetric screens have been applied have primarily been performed in *E. coli*, which naturally utilizes the MEP pathway. Gallagher et al. (2003) screened a cDNA library constructed from poly(A)⁺ maize endosperm RNA in *E. coli* strains heterologously producing lycopene and β -

carotene, respectively. They hereby identified an IPP isomerase encoded by a novel gene which could improve carotenoid production. The isomerization of IPP to DMAPP can therefore be concluded to be a limiting step in isoprenoid synthesis in *E. coli*. A similar approach was utilized by Kajiwarra et al. (1997), who isolated IPP isomerase genes from *Xanthophyllomyces dendrorhous* and *Haematococcus* cDNA libraries, using a screen based on a β -carotene producing *E. coli* strain. Kang et al. (2005) constructed a gDNA library from *E. coli* and used it to transform a lycopene producing *E. coli* strain. The plasmid used to transform *E. coli* making it carotenogenic contained in addition to the lycopene biosynthetic genes also the *ipi* gene from *H. pluvialis*, which encodes a bacterial IPP isomerase, in order to alleviate this limiting step hereby allowing identification of the next limiting step in isoprenoid biosynthesis. They found that plasmids from their library harboring the following genes improved lycopene production: *crl*, *rpoS*, *appY*, and *dxs*. Of these genes, *dxs* which encode the first enzyme of the MEP pathway, could be directly linked to isoprenoid production, whereas the others encode proteins with more global functions such as RNA polymerase and regulatory proteins. By applying a combinatorial expression approach the authors found that the combination of *dxs* and *appY* resulted in the highest production of lycopene (4.7 mg/gDCW) which corresponds to an approximate 8 fold increase compared to the parental strain. The finding that the overexpression of *idi* and *dxs* can improve isoprenoid production corresponds well with previous studies by Matthews and Wurtzel (2000) and Kim and Keasling (2001). Jin and Stephanopoulos (2007) utilized an approach similar to that of Kang et al. (2005), but instead of only screening the library in a single strain they screened their gDNA library in three strains with different genetic backgrounds. These strains were DH5 α , WS140 (*dxs*, *idi*, *ispFD* chromosomal overexpression – Yuan et al. 2006) and WS140 Δ *gdh* Δ *ace* Δ *fdhF* (triple knockout strain constructed by Alper et al., 2005a from knockout targets identified through stoichiometric flux balance analysis, FBA). In the plasmids isolated from the gDNA library screened in DH5 α the most frequently identified genes were the *dxs* and *idi*, in accordance with prior findings. In strain WS140, where the *dxs* and *idi* genes were overexpressed, the *rpoS* gene was the one that most frequently came up in the screen. *rpoS* was also identified in a single plasmid isolated from the DH5 α screen. In the triple knockout background, *rpoS* was also the gene identified most frequently. These articles demonstrate how the combination of DNA libraries and colorimetric screening can be applied in an iterative manner to find novel targets one by one, hereby gradually alleviating limiting steps and improving isoprenoid production.

Not only overexpression targets are of interest in the effort towards construction of microbial cell factories; also deletion targets have been a preferred approach. Knockout targets can be identified through the utilization of random transposon mutagenesis. Tao et al. (2005) initially aimed at using lycopene as a screen-able reporter for copy number of a ColE1-type plasmid carrying the lycopene biosynthetic genes. Besides identifying gene knockouts affecting plasmid copy number, they also identified targets that by other mechanisms altered lycopene production. None of these were however linkable to isoprenoid production in any obvious way. However, one of the mutants with increased lycopene production had a transposon insert within the *hscB*

gene, which encodes a chaperone involved in iron-sulfur cluster biogenesis. This is opposed to what might have been expected, since the enzymes catalyzing the last two steps of the MEP pathway (IspG and IspH) are iron sulfur cluster proteins. Alper et al. (2005b) also used transposon mutagenesis but with the aim to create isoprenoid overproducing knockout mutants. The mutagenesis was performed on a strain with modifications similar to those of WS140. From this search three transposon inserts were obtained, whereof two were inserted in ORFs (*rssB* and *yjfP*) and therefore can be considered knockouts. The third was inserted between the *yjiD* and its promoter. All combinations of the three mutations were constructed and combined with the targets identified by FBA [Alper et al., 2005a]. Interestingly two global maxima for lycopene production were identified; the first being a combination identified solely through FBA and the other being the combination of the transposon modified *yjiD* gene and $\Delta gdhA$ and $\Delta aceE$. *yjiD* was also identified in a single plasmid isolated from the gDNA library screen by Jin et al. (2007) in the WS140 $\Delta gdh\Delta ace\Delta fdhF$ background strain. This corroborates the beneficial effect of the combination of these three mutations and suggests that insertion of the transposon between the *yjiD* gene and its promoter resulted in overexpression of this gene.

Deletions and overexpression constructs represent metabolic extremes, and it is unlikely that these will constitute the genotypic basis leading to an optimal production phenotype. However, the simultaneous modulation of multiple genes is troublesome and hampered by the lack of sufficient molecular biology techniques. To remedy this limitation Alper et al. (2007) invented a technique which they coined global transcription machinery engineering (gTME). gTME allowed them to simultaneously modulate all genes for which transcription was mediated through global transcription factors, hereby altering the transcription profile of the cell. However, a large number of cells harboring different transcription factor mutants have to be screened since an enormous number of combinations can be achieved already within each of the genes. As a proof of concept they utilized the improvement of lycopene production in *E. coli*, since this phenotype had been found to be very well suited for high-throughput screening. Using gTME in conjunction with colorimetric screening they could further improve upon their best deletion mutants described above.

Although carotenoids have also been synthesized in *Saccharomyces cerevisiae* [Yamano et al. 1994]; we have not found any example in the literature describing colorimetric screening based on heterologously production of carotenoids in this organism. Engineering of *S. cerevisiae* for improved isoprenoid production has rather relied on rational strategies. However, improvement of carotenoid production within the natural carotenoid producing yeast *Xanthophyllomyces dendrorhous* has been achieved through colorimetric screening. These screens required the addition of inhibitors of carotenoid biosynthesis to allow for discrimination between wild-type and mutant colonies [Lewis et al. 1990; Chumpolkulwong et al. 1997]. Comparable carotenoid yields have been achieved in *S. cerevisiae* and *X. dendrorhous*, but colorimetric saturation and the requirement for inhibitory compounds could perhaps be a reason that colorimetric screens have not been applied for *S. cerevisiae*. Also other yeasts, such as *Candida utilis*, have been

engineered for production of carotenoids [Miura et al. 1998], and with the development of robust carotenoid based colorimetric screens designed for yeasts the potential of these organisms as hosts for heterologous isoprenoid production could be greatly improved and developed into a good alternative to *E. coli*. Figure 1.1 shows *S. cerevisiae* producing lycopene, β -carotene, and astaxanthin, respectively.

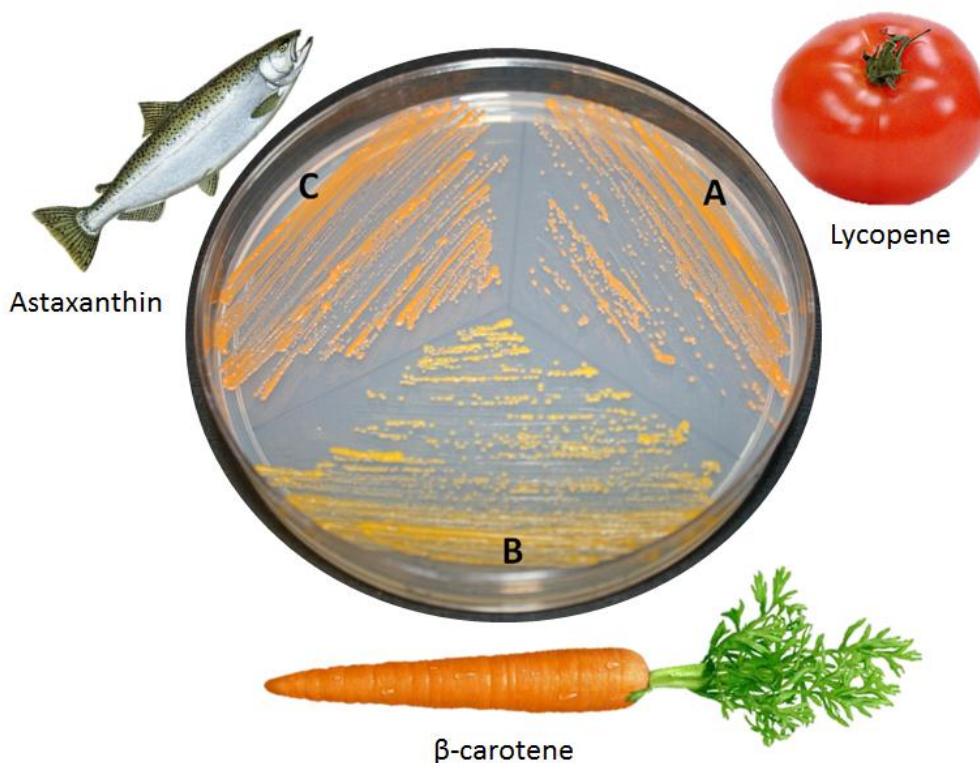


Figure 1.1: *S. cerevisiae* strains engineered for production of different carotenoids. The food source primarily associated with the carotenoid being produced by each engineered yeast is presented next to the strain producing it. Prevalent carotenoid being produced is A: lycopene, B: β -carotene, and C: astaxanthin.

1.4 Polyketides and flavonoids

Colorimetric screening of surrogate molecules has also been applied for these two groups of metabolites, but less extensively than for the isoprenoids. Malonyl-CoA is a precursor molecule applied in biosynthesis of both polyketides and flavonoids, and its abundance within microbial cells is important for their high level production. Several rational engineering strategies have increased malonyl-CoA availability [Leonard et al. 2008], but no combinatorial approach has to our knowledge been applied. However, Zha et al. (2008) utilized directed evolution to optimize *Pseudomonas fluorescens* phloroglucionol synthase heterologously expressed in *E. coli* by screening for colonies exhibiting purple halos when grown on agar plates containing Gibb's reagent (*N*,2,6-trichloro-*p*-benzoquinoneimine). Color formation was a result of the reaction of the polyketide phloroglucinol synthesized from three molecules of malonyl-CoA and Gibb's reagent. Colonies selected from the initial screen were further tested in 96-well plates by the reaction between the phloroglucinol in a cell-free supernatant and cinnamaldehyde. This method allowed the authors to isolate several genes encoding mutated phloroglucionol synthases with improved catalytic properties. This screening method in conjunction with one of the mutated phloroglucionol synthases has the potential to be converted into a more general screen for malonyl-CoA precursor availability and hereby to be applied towards improving heterologous production of polyketides and flavonoids.

Another key precursor metabolite for synthesis of flavonoids is L-tyrosine, and screening methods for detecting microbes with improved L-tyrosine production has been developed. The first one was as proof of concept demonstrated using *E. coli*; Lütke-Eversloh and Stephanopoulos (2007) demonstrated that L-tyrosine secreted to the fermentation broth of micro-titer plate cultures could be quantified by derivatization with 1-nitroso-2-naphthol. This method had previously been demonstrated to permit colorimetric and fluorometric determination of L-tyrosine [Udenfriend and Cooper 1952; Waalkes and Udenfriend 1957] but was adapted for high-throughput assessment of L-tyrosine production. By using fluorescence measurements Lütke-Eversloh and Stephanopoulos circumvented the tedious work of removing the cells from the supernatant in many samples prior to L-tyrosine quantification and simplified the procedure from ten steps to four steps, even further improving its applicability for high-throughput screening. Santos and Stephanopoulos (2008) invented another colorimetric method for screening *E. coli* mutants for improved L-tyrosine production. By expression of a tyrosinase encoded by the *mela* gene from *Rhizobium etli* they converted the L-tyrosine into 3,4-dihydroxy-L-phenylalanine (L-DOPA) and further by oxidation into dopachrome, which can undergo non-enzymatic polymerization to form the black pigment melanin. After optimizing the composition of the growth medium to maximize melanin production they could use the black melanin phenotype to screen a transposon mutagenesis library for *E. coli* colonies on agar plates for L-tyrosine overproducers. This method is simpler than the aforementioned method since less pipetting and handling is involved, and pseudo-high-throughput can be achieved without the aid of robotics.

These examples show how microbes may be optimized for production of some of the key precursors for flavonoid biosynthesis, but the pathway downstream thereof also deserves optimization to obtain efficient production strains. Whereas most heterologous flavonoid production strategies had relied on phenylpropanoic precursor supplementation, Santos et al. (2011) demonstrated the synthesis of the main flavonoid precursor naringenin from glucose. Their engineered strain permits investigations towards identifying both local and global targets that can further improve flavonoid precursor production. However, an efficient screen is required, which could be performed if naringenin could be converted to a metabolite resulting in a quantifiable colored phenotype. Fortunately several flavonoid compounds have been found to exhibit properties making them ideal as colorants [Winkel-Shirley 2001]. One sub-group of these metabolites is the anthocyanins, which have red or blue colors. Anthocyanin has been applied to assess the heterogeneity of cultured strawberry cells. This was done by image analysis which was used to analyze microscope pictures for red coloration by comparing the RGB-color parameters of the images [Miyanaga et al. 2000]. Anthocyanins have been heterologously produced in *E. coli* by Yan et al. (2005). This was done by expressing four genes encoding the following enzymes: 3 β -hydroxylase from *Malus domestica*, dihydroflavonol 4-reductase from *Anthurium andraeanum*, anthocyanidin synthase also from *M. domestica*, and UDP-glucose:flavonoid 3-*O*-glucosyltransferase from *Petunia*. If the genes encoding these four enzymes were expressed in the strain created by Santos et al. (2011) it should convert the colorless naringenin into the colored anthocyanins which in a manner similar to the application of lycopene described above could be used for optimizing *E. coli* for heterologous production of flavonoids. Since all the genes encoding the pathway to anthocyanins are known it should further be possible to use this approach to identify novel targets for improving flavonoid production in other microbes.

1.5 Tools for improving throughput

Even though visual inspection of microbial colonies grown on agar plates is an established and efficient screening method, it has drawbacks. Humans have a good visual perception, but details such as color intensities relative to other objects is not clearly imbedded in the memory, a fact that often forces the investigator to refer back and forth between the individual objects in order to dissociate objects with certain traits, making this approach time consuming [Yang and Youvan 1988]. Furthermore, different persons may have different color perception. To remedy these drawbacks Digital Imaging Spectroscopy (DIS) was invented. DIS enables the researcher to perform multiple parallel spectroscopic assessments of 2D images. The main components of DIS are a digital camera, light source, and an array of light filters. With DIS the wavelength of the light source and the wavelength of the light recorded by the digital camera can be controlled, allowing for the generation of multiple spectra. DIS can be used to obtain both reflectance and absorption spectra for the target. Spectra can be obtained in the range 400-900 nm (visible to near infrared), to hereby include the assessment of features not visible to the human eye. Furthermore, DIS can be applied to obtain fluorescence spectra. With DIS, up to 25 agar plates can be processed simultaneously containing up to 500 colonies each. For an introduction to DIS the reader is referred to Youvan et al. (1995). DIS has, e.g. been used for directed evolution of galactose oxidase [Delagrave et al. 2001], investigation of the effects of mutations on light-harvesting complexes of the photosynthetic bacterium *Rhodobacter capsulatus* [Arkin et al. 1990], and the assessment of the reflectance spectrum of a red yeast [Yang 1994]. DIS could be applied to chromogenic screens to improve the throughput but has not become a widespread method.

Digital cameras, scanners and personal computers are common items found in most laboratories. This allows for less expensive and more readily available methods for automating the colorimetric screening of microbial populations on solid surfaces as compared to DIS. Gallagher et al. (2003) applied a Hewlett Packard Scan Jet 6100C scanner to obtain images of their colonies of carotenogenic *E. coli* cells transformed with a maize endosperm cDNA library. The Images were processed using Adobe Photoshop 4.0.1. This is an example of application of “everyday” lab equipment to evaluate colony coloration caused by production of a heterologous compound.

We followed a similar approach when we developed a method for quantifying colony coloration of *S. cerevisiae* engineered to produce lycopene. The method was based on the acquirement of digital images of carotenogenic yeast cells and processing of these images using the image analysis software ImageJ. By constructing a *S. cerevisiae* strain with the lycopene biosynthetic pathway under control of the copper inducible *CUP1* promoter we could correlate the processed colony coloration intensities for yeast cells grown on a certain copper concentration with the measured lycopene production in liquid cultures at the same copper concentrations [Carlsen et al.

201Xa]. This method offers a simple approach to improving colorimetric screening of microbial populations expressing red pigment molecules.

Dr. K.R. Patil and Dr. A.E. Lantz have constructed a setup which allows for measuring growth rates and pigment production of microbial cultures grown in micro-titer plates hereby allowing for the selection of mutants with the highest productivity [personal communication]. The system has as proof-of-concept been tested on heterologous lycopene production by *S. cerevisiae* and actinorhodin production by *Streptomyces coelicolor*. The setup is composed of two flat-bed scanners placed under an orbital shaker located within an incubator that offers control of temperature and humidity combined with the blockage of any interfering light sources. The entire setup is controlled by a computer which at set time points stops the orbital shaker and acquires scans of the bottom of each micro-titer plate in which microbes are growing. The micro-titer plate images are analyzed using a custom designed software that calculates product formation and growth measurements.

Although the aforementioned methods greatly improve colorimetric screening, they can only be considered as a step towards the setup of real high-throughput methods. Handling and picking of individual colonies are so far still performed manually. If the colorimetric screening methods were to be interfaced with robotic colony pickers for transferring the mutants identified to exhibit a desired trait into micro-titer plates, combined with micro-titer-plate handling and automation for further analysis, they would represent true high-throughput methods [Youvan et al. 1995]. The technologies for doing this are definitely available. For further descriptions of screening automation the reader is referred to Hamilton (2002).

1.6 Conclusions and future perspectives

A wealth of colored metabolites is produced in nature. They represent a great potential since they can be heterologously biosynthesized and hereby be utilized as surrogate molecules in the development of novel colorimetric screens for improving microbial production. Ideal surrogate molecules and their biosynthetic pathways should meet as many of the following criteria as possible:

- A short pathway from precursor metabolite to reporter metabolite
- Proportional response in reporter metabolite concentration to changes in precursor metabolite concentration
- Balanced pathway to avoid buildup of intermediates between precursor and reporter metabolite
- Easily quantifiable compound to ensure high throughput
- High stability and molecular integrity
- Similar physical and chemical properties to the metabolite for which it is used as a surrogate
- No or low toxicity to the chosen production organism
- Impose small metabolic burden
- High signal to noise ratio

Most groups of metabolites span a wide range within properties such as hydrophobicity and volatility. This can create problems in the use of surrogate compounds. E.g. since lycopene is very hydrophobic and accumulated inside cells when heterologously produced in microorganisms [Yamano et al., 1994], its application as surrogate compound could result in the isolation of mutants that influence intracellular storage capacity and would thus not benefit the production of excreted isoprenoid metabolites. Another issue in the application of heterologous metabolites in surrogate screens is the host response to these. Verwaal et al. (2010) found that heterologous production of carotenoids in *S. cerevisiae* triggers the pleiotropic drug resistance genes, suggesting that these compounds have a detrimental effect on the yeast cells. When carotenoid molecules are applied in a *S. cerevisiae* screening endeavor, there is therefore a possibility that some of the recovered genotypes are involved in increasing the tolerance towards them. If the same is valid for *E. coli*, this could be the reason that some of the recovered genotypes in the lycopene based screens are not directly linked to improving precursor availability. Indeed, when we improved heterologous production of squalene, which has a molecular structure very similar to lycopene, in *E. coli*, we observed that it had a negative effect on the growth of the cells [Carlsen et al. 201Xb]. This supports the case of *E. coli* being affected by the expression of hydrophobic isoprenoids. Even though surrogate compounds have to be selected with care, their application in microbial screens have helped enormously in the effort towards isolation of novel genotypes for improving microbes for production of natural products.

Due to current sub-optimal cost-effectiveness of microbial natural product production, only high value natural products or molecules that cannot be obtained by other means than microbial production can so far be cost-competitively biosynthesized. With more culturable microorganisms exhibiting novel properties being identified and genetic and molecular biology techniques being developed to engineer these, the strategy of applying microbial production platforms for production of less costly natural products is not far from being realizable; the host organism for the production of each individual group of metabolites just has to be carefully selected based on the properties of the metabolites [Zhang et al. 2011]. After choosing a microorganism as starting point, one needs to optimize it for production of the given compound. In this endeavor, colorimetric screening will play an important role with a development of novel screens concurrently with automation of already established screens, since most of these screens are applicable for multiple species of microbes. Metabolites for which screening methods exist are still greatly exceeded by the methods available to manipulate microorganisms, so the development of screening methods based on surrogate compounds is definitely an area that will gain increased attention in the future.

1.7 Competing interests

The authors declare that they have no competing interest.

1.8 References

Ajikumar PK, Tyo K et al (2008) Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol Pharm* 5(2): 167–190

Alper H, Jin Y-S et al (2005a) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng* 7:155-164

Alper H, Miyaoku K et al (2005b) Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* 23(5):612-616

Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng* 9:258-267

Arkin AP, Goldman ER et al (1990) Applications of imaging spectroscopy in molecular biology II. Colony screening based on absorption spectra. *Biotechnology (N Y)* 8(8):746-749

Carlsen S, Nielsen ML et al (201Xa) The application of image analysis for semi-high-throughput screening for isopentenyl pyrophosphate overproducing *Saccharomyces cerevisiae* colonies through utilization of the surrogate isoprenoid pigment lycopene.

Carlsen S, Zhang H et al (201Xb) Engineering *Escherichia coli* for production of squalene: utilization of the *ERG9* gene from *Saccharomyces cerevisiae*.

Chiang S-J (2004) Strain improvement for fermentation and biocatalysis processes by genetic engineering technology. *J Ind Microbiol Biot* 31:99-108

Chatterjee R, Yuan L (2006) Directed evolution of metabolic pathways. *Trends Biotechnol* 24(1):28-38.

Chumpolkulwong N, Kakizono T et al (1997) Increased astaxanthin production by *Phaffia rhodozyma* mutants isolated as resistant to diphenylamine. J Ferment Bioeng 83(5):429-434

Delagrave S, Murphy DJ (2001) Application of a very high-throughput digital imaging screen to evolve the enzyme galactose oxidase. Protein Eng 14(4):261-267

Dove A (2003) Screening for content — the evolution of high throughput. Nat Biotechnol 21(8):859-864

Gallagher CE, Cervantes-Cervantes M et al (2003) Surrogate biochemistry: use of *Escherichia coli* to identify plant cDNAs that impact metabolic engineering of carotenoid accumulation. Appl Microbiol Biotechnol 60:713-719

Harvey AL (2007) Natural products as a screening resource. Curr Opin Chem Biol 11:480-484

Hamilton S (2002) Introduction to screening automation. Method Mol Biol 190:169-193

Jin Y-S, Stephanopoulos G (2007) Multi-dimensional gene target search for improving lycopene biosynthesis in *Escherichia coli*. Metab Eng 9:337-347

Kajiwarra S, Fraser PD et al (1997) Expression of an exogenous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. Biochem J. 324:421-426

Kang MJ, Lee YM et al (2005) Identification of genes affecting lycopene accumulation in *Escherichia coli* using a shot-gun method. Biotechnol Bioeng 91(5):636-642

Kim SW, Keasling JD (2001) Metabolic engineering of the non-mevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. Biotechnol Bioeng 72:408-415

Lee PC, Petri R et al (2005) Directed evolution of *Escherichia coli* farnesyl diphosphate synthase (IspA) reveals novel structural determinants of chain length specificity. *Metab Eng* 7:18-26

Leonard E, Ajikumar PK et al (2010) Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc Natl Acad Sci USA* 107(31):13654-13659

Leonard E, Yan Y et al (2008) Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. *Mol Pharm* 5(2):257-265

Lewis MJ, Ragot N et al (1990) Selection of astaxanthin-overproducing mutants of *Phaffia rhodozyma* with β -Ionone. *Appl Environ Microbiol* 56(9):2944-2945

Lütke-Eversloh T, Stephanopoulos G (2007) A semi-quantitative high-throughput screening method for microbial L-tyrosine production in microtiter plates. *J Ind Microbiol Biotechnol* 34:807-811

Matthews PD, Wurtzel ET (2000) Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. *Appl Microbiol Biotechnol* 53:396-400

Meyer H-P, Turner NJ (2009) Biotechnological manufacturing options for organic chemistry. *Mini-Rev Org Chem* 6:300-306

Mills DR, Peterson RL et al (1967) An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc Natl Acad Sci USA* 58(1):217-224

Miura Y, Kondo K et al. (1998) Production of the carotenoids lycopene, β -carotene, and astaxanthin in the food yeast *Candida utilis*. *Appl Environ Microbiol* 64:1226-1229.

Miyanaga K, Seki M et al (2000) Quantitative determination of cultured strawberry-cell heterogeneity by image analysis: effects of medium modification on anthocyanin accumulation. *Biochem Eng J* 5:201-207

Octave S, Thomas D (2009) Biorefinery: toward an industrial metabolism. *Biochimie* 91:659-664

Ohnuma S, Nakazawa T et al (1996) Conversion from farnesyl diphosphate synthase to geranylgeranyl diphosphate synthase by random chemical mutagenesis. *J Biol Chem* 271(17):10087-10095

Ohnuma S, Suzuki M et al (1994) Archaeobacterial ether-linked lipid biosynthetic gene. Expression cloning, sequencing, and characterization of geranylgeranyl-diphosphate synthase. *J Biol Chem* 269(20):14792-14797

Olsen M, Iverson B et al (2000) High-throughput screening of enzymes libraries. *Curr Opin Biotechnol* 11:331-337

Parekh S, Vinci VA et al (2000) Improvement of microbial strains and fermentation processes. *Appl Microbiol Biotechnol* 54:287-301

Reymond J-L, Babiak P (2007) Screening systems. *Adv Biochem Engin/Biotechnol* 105:31-58

Rowlands RT (1984) Industrial strain improvement: mutagenesis and random screening procedures. *Enzyme Microb Technol* 6:3-10

Santos CNS, Koffas M et al (2011) Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab Eng* 13:392-400

Santos CNS, Stephanopoulos G (2008) Melanin-based high-throughput screen for L-tyrosine production in *Escherichia coli*. *Appl Environ Microbiol* 74(4):1190-1197

Tao L, Jackson RE et al (2005) Isolation of chromosomal mutations that affect carotenoid production in *Escherichia coli*: mutations alter copy number of ColE1-type plasmids. *FEMS Microbiol Lett* 243:227-233

Tao L, Wilczek J et al (2006) Engineering a β -carotene ketolase for astaxanthin production. *Metab Eng* 8:523-531

Tyo KE, Alper HS et al (2007) Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol* 25(3):132-137.

Udenfriend S, Cooper JR (1952) The chemical estimation of tyrosine and tyramine. *J Biol Chem* 196:227-233

Umeno D, Tobias AV et al (2005) Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiol Mol Biol R* 69(1):51-78

Verwaal R, Jiang Y (2010) Heterologous carotenoid production in *Saccharomyces cerevisiae* induces the pleiotropic drug resistance stress response. *Yeast* 27(12):983-998

Waalkes TP, Udenfriend S (1957) A fluorometric method for the estimation of tyrosine in plasma and tissues. *J Lan Clin Med* 50:733-736

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol* 126: 485-493

Yamano S, Ishii T et al (1994) Metabolic engineering for production of β -carotene and lycopene in *Saccharomyces cerevisiae*. Biosci Biotech Biochem 58:1112-1114

Yan Y, Chemler et al (2005) Metabolic engineering of anthocyanin biosynthesis in *Escherichia coli*. Appl Environ Microbiol 71(7):3617-3623

Yang MM (1994) Digital imaging spectroscopy of microbial colonies. Am Biotechnol Lab 12(6):18-20

Yang MM, Youvan DC (1988) Applications of Imaging Spectroscopy in Molecular Biology: I. Screening Photosynthetic Bacteria. Nat Biotechnol 6:939-942

Youvan DC, Goldman E (1995) Digital imaging spectroscopy for massively parallel screening of mutants. Method Enzymol 246:732-748

Yuan LZ, Rouviere PE et al (2006) Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *Escherichia coli*. Metabol Eng 8:79-90

Zha W, Rubin-Pitel SB et al (2008) Exploiting genetic diversity by directed evolution: molecular breeding of type III polyketide synthases improves productivity. Mol Biosyst 4:246-248

Zhang H, Boghigian BA et al (2011) Methods and options for heterologous production of complex natural products. Nat Prod Rep 28:125-151

Zhao H, Chockalingam K et al (2002) Directed evolution of enzymes and pathways for industrial biocatalysis. Curr Opin Biotechnol 13:104-110

Chapter 2 – Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

Simon Carlsen*†, Parayil Kumaran Ajikumar†, Luca Riccardo Formenti*, Kang Zhou†‡, Too Heng Phon‡, Michael Lynge Nielsen*¹, Anna Eliasson Lantz*, Morten C. Kielland-Brandt*, Gregory Stephanopoulos†[§]

[§] Corresponding author.

* Department of Systems Biology, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

† Department of Chemical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA.

‡ Chemical and Pharmaceutical Engineering Program, Singapore-MIT Alliance, 117546 Singapore.

¹ Present address: Novozymes A/S, Krogshoejvej 36, DK-2880 Bagsvaerd, Denmark.

Manuscript intended for submission to Metabolic Engineering

2.1 Abstract

Transfer of a biosynthetic pathway between evolutionary distant organisms can create a metabolic shunt capable of bypassing the native regulation of the host organism, hereby improving the production of secondary metabolite precursor molecules of important natural products. In most cases transfer of the genes encoding the pathway enzymes followed by balancing of the individual steps is sufficient to achieve expression of a fully functional heterologous pathway. In this study we integrated the *Escherichia coli* genes encoding the 2-*C*-methyl-D-erythritol 4-phosphate (MEP)-pathway into the genome of *Saccharomyces cerevisiae* with the aim of bypassing regulation of the mevalonate (MVA)-pathway and create an improved microbial platform for production of isoprenoids. We found that most of the pathway is active, synthesizing 2-*C*-methyl-D-erythritol-2,4-cyclodiphosphate (MEC) but appears to lack functionality of the last two steps of the MEP-pathway, catalyzed by the [4Fe-4S] iron sulfur cluster proteins encoded by *ispG* and *ispH*. We detected production of not only MEC but also deoxyxylulose 5-phosphate in *S. cerevisiae*, which has never before been reported. To try to functionalize the last two steps of the MEP-pathway we co-expressed the genes for the *E. coli* iron sulfur cluster (ISC)-assembly machinery. By deleting *ERG13*, thereby incapacitating the MVA-pathway, in conjunction with labeling experiments with U-¹³C₆ glucose and growth experiments, we found that the ISC-assembly machinery was unable to functionalize *ispG* and *ispH*. However, we have found that *leuC* and *leuD*, encoding the hetero-dimeric iron sulfur cluster protein isopropylmalate isomerase, can complement the *S. cerevisiae leu1* auxotrophy. This is to our knowledge the first time a bacterial iron-sulfur cluster protein has been functionally expressed in the cytosol of *S. cerevisiae* and shows that *S. cerevisiae* has the capability to functionally express at least some bacterial iron sulfur cluster proteins in its cytosol.

2.2 Introduction

Isoprenoids, comprising more than 55,000 compounds, holds a great potential for the discovery of novel pharmaceuticals, flavors, fragrances, and in general high-value chemicals. The complexity of these compounds and their scarce availability in nature however render their production highly costly and difficult via chemical synthesis or extraction, hereby leaving heterologous biosynthesis in organisms amenable to genetic engineering the most obvious strategy of production. However, biosynthesis is not straightforward and requires optimization to establish an effective downstream pathway for obtaining the product of interest, avoid byproduct formation, and improve precursor supply. In cases where it is impossible to establish a total biosynthetic scheme due to, e.g., the genes encoding the required enzymatic steps being unknown, the cells being unable to functionally express the required genes, or the intermediates of the pathway being toxic to the host cell, it is advantageous to use a combinatorial approach comprising a fermentation step providing the isoprenoid backbone followed by a chemical step converting the backbone to the product of interest. Nevertheless, for any of the aforementioned approaches to be viable a high flux through the upstream pathway is required to ensure a high isoprenoid precursor supply [Maimone and Baran, 2007; Ajikumar et al. 2008].

All isoprenoids are biosynthesized from the C-5 precursor isopentenyl diphosphate (IPP), which can be produced through two distinct pathways, namely the mevalonate (MVA) pathway from acetyl-CoA and the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway from pyruvate and glyceraldehyde-3-phosphate (G3P) (Figure 2.1). Some organisms like plants utilize both pathways, but most organisms utilize only one of the two, hereby opening an approach to bypass the native regulation of the production organism of choice by expression of the heterologous pathway [Martin et al., 2003; Maury et al., 2008]. The first example by Martin et al. (2003) showed this strategy to be an effective way of improving the production of isoprenoids. The MVA-pathway was engineered into *Escherichia coli*, which naturally only possesses the MEP-pathway. In order to do this, two operons were created based on 4 codon-optimized genes each, encompassing upstream and downstream parts of the mevalonate pathway, respectively. By co-expressing the two plasmids in an *E. coli* strain harboring an *ispC* deletion, which causes a requirement for deoxyxylulose 5-phosphate (DXP), Martin et al. [2003] were able to obtain DXP-non-requiring cells, hereby demonstrating that the heterologous MVA-pathway can provide an amount of IPP sufficient to sustain growth. They further showed that the engineered *E. coli* strain could be utilized to produce high amounts of the artemisinin precursor amorpha-4,11-diene, and they hereby also demonstrated proof-of-concept of generating a microbial platform engineered to produce high amounts of IPP which can be converted into an isoprenoid of interest for which the synthase gene(s) are known.

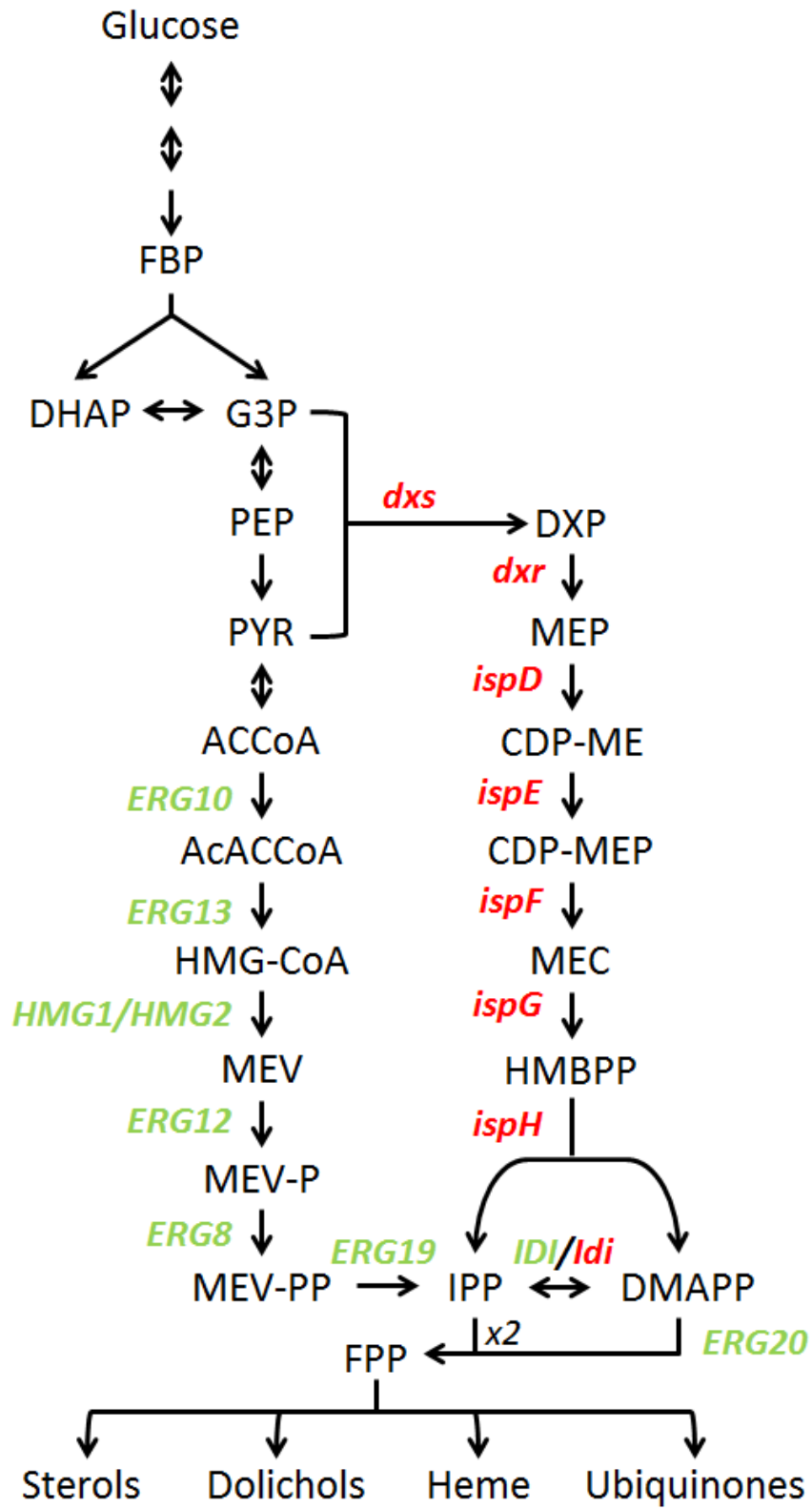


Figure 2.1: Overview of isoprenoids biosynthetic pathways. Green gene names are mevalonate pathway genes and red gene names are 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway genes. Gene / protein: *ERG10* / acetoacetyl-CoA thiolase, *ERG13* / 3-hydroxy-3-methyl-glutaryl-CoA synthase, *HMG1* and *HMG2* / 3-hydroxy-3-methyl-glutaryl-CoA reductase, *ERG12* / mevalonate kinase, *ERG8* / phosphomevalonate kinase, *ERG19* / mevalonate-5-diphosphate decarboxylase, *IDI* and *idi* / isopentenyl diphosphate isomerase, *ERG20* / Farnesyl diphosphate synthase, *dxs* / 1-deoxy-D-xylulose-5-phosphate synthase, *dxr* / 1-deoxy-D-xylulose-5-phosphate reductase, *ispD* / 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase, *ispE* / 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, *ispF* / 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, *ispG* / 1-hydroxy-2-methyl-2-(E)-butenyl-4-phosphate synthase, *ispH* / 1-hydroxy-2-methyl-2-(E)-butenyl-4-phosphate reductase. Pathway intermediates: FBP – fructose biphosphate, DHAP – dihydroxyacetone phosphate, G3P – glyceraldehyde 3-phosphate, PEP – phosphoenolpyruvate, PYR – pyruvate, ACCoA – acetyl-CoA, AcACCoA – acetoacetyl-CoA, HMG-CoA – 3-hydroxy-3-methyl-glutaryl-CoA, MEV – mevalonate, MEV-P – mevalonate-5-phosphate, MEV-PP – mevalonate diphosphate, IPP – isopentenyl diphosphate, DMAPP – dimethylallyl diphosphate, FPP – farnesyl diphosphate, DXP – 1-deoxy-D-xylulose-5-phosphate, MEP – 2-C-methyl-D-erythritol-4-phosphate, CDP-ME – 4-diphosphocytidyl-2-C-methyl-D-erythritol, CDP-MEP – 2-phospho-4-diphosphocytidyl-2-C-methyl-D-erythritol, MEC – 2-C-methyl-D-erythritol-2,4-cyclodiphosphate, HMBP – 1-hydroxy-2-methyl-2-(E)-butenyl-4-phosphate.

Expression of a non-native, multi-step pathway yields a deregulation of the flux, which thus increases desirably, but this also implies that the pathway cannot self-regulate and therefore in most cases has to be additionally engineered to avoid bottlenecks which would otherwise cause suboptimal pathway performance. This was indeed the case for the strain engineered by Martin et al. [2003]. When the authors analyzed their strain by liquid chromatography – mass spectrometry (LC-MS), they found a buildup of the MVA-pathway intermediate 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), which inhibited cell growth. The buildup was attributed to HMG-CoA reductase being inefficient at converting HMG-CoA to MVA. The problem was alleviated by modulation of the expression level of the *tHMGR* gene, encoding a truncated HMG-CoA reductase, hereby demonstrating the importance of balancing the carbon flux of a heterologously expressed multi-step pathway [Pitera et al. 2007].

Saccharomyces cerevisiae, on the other hand, naturally utilizes the MVA-pathway for production of isoprenoids, so in order to engineer a *S. cerevisiae* platform for production of isoprenoids [Maury et al. 2008] expressed the seven genes encompassing the mevalonate pathway on two 2 μ self-replicating plasmids together with a valencene synthase gene. The MEP-pathway has been found to be stoichiometrically superior to the MVA-pathway [Dugar and Stephanopoulos, 2011], making it obvious to try this pathway for heterologous expression. Maury and coworkers amplified the seven MEP-pathway genes by PCR from *E. coli* genomic DNA and expressed them using a dual galactose inducible promoter system based on the *GAL1*-promoter and the *GAL10*-promoter. To verify that the MEP-pathway was active they inhibited the MVA-pathway

with the hypocholesterolemic drug lovastatin, which specifically inhibits the HMG-CoA reductase [Alberts et al., 1980], and since the MVA pathway supplies precursors for ergosterol and other essential compounds, this drug inhibits growth of wild type yeast. The MEP pathway appeared to support growth, thwarting the lovastatin mediated inhibition, hereby indicating that the heterologous expressed MEP-pathway is active. This work led to the patenting of MEP-pathway transformed microorganism for production of isoprenoids [Maury et al., 2008].

So far the aforementioned examples are the only two examples of transferring the complementary IPP biosynthetic pathway into an organism devoid of either the MVA-pathway or the MEP-pathway. Transferring an entire pathway is a tedious and challenging task, which is probably why it have only been done in the model organisms *E. coli* and *S. cerevisiae*, for which well-developed genetic engineering tools and expression systems are available. In this study we likewise decided to engineer the MEP-pathway into *S. cerevisiae*, but to create a stable strain amenable for further genetic modification we chose to chromosomally integrate the seven MEP-pathway genes from *E. coli* with the addition of the *Idi* gene, which encodes the *E. coli* IPP/dimethylallyl diphosphate (DMAPP) isomerase. These eight genes were obtained synthetically and codon optimized to improve their expression in *S. cerevisiae* to compensate for the lower expression level caused by having a single copy of the gene in the chromosome compared to the 2 μ -based multi-copy plasmids. To test whether the chromosomally integrated pathway were functional expressed we decided to delete the *ERG13* gene, which encodes the HMG-CoA synthase of the MVA-pathway. Contrarily to what was presented by Maury et al. [2008], the *E. coli* MEP-pathway chromosomally expressed in *S. cerevisiae* was unable to support growth in the absence of IPP supplied by the MVA-pathway. To investigate what caused the MEP-pathway to be inactive, and where in the pathway the problem occurred, we analyzed the strain for MEP-pathway metabolites using liquid chromatography – mass spectrometry. We detected both DXP and 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEC). The results indicate that the MEP pathway is active until MEC, and the problem thus originates from lack of sufficient expression of either of the two genes (*ispG* and *ispH*) downstream of MEC or both. We hereafter searched the literature to elucidate whether *S. cerevisiae* meets the co-factor requirements for the entire MEP-pathway, cf. Table 2.1. From Table 2.1 it can be seen that all co-factor requirements for the first five pathway steps are fulfilled. However, step 6 and 7 requires ferredoxin or flavodoxin and ferredoxin (flavodoxin) NADP⁺ reductase to couple the reaction to NADPH for electron transfer. Furthermore, both *ispG* and *ispH* are iron-sulfur cluster proteins which need to be loaded with a [4Fe-4S]-iron-sulfur cluster (ISC) to be functional [Seemann and Rohmer, 2007]. Ferredoxin is also an [2Fe-2S] ISC-protein [Blaschkowski et al. 1981].

Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

Table 2.1: MEP-pathway reactions and cofactors

Protein name/ gene name(s)	Substrate(s)/product(s)	Cofactor(s)	Reference	EcoCyc accession number
Dxs/ <i>dxs</i>	pyruvate + D-glyceraldehyde-3-phosphate + H ⁺ / CO ₂ + 1-deoxy-D-xylulose 5-phosphate	Thiamine diphosphate; Mg ²⁺ ; Mn ²⁺	Xiang et al., 2007	G6237
Dxr/ <i>dxr</i>	1-deoxy-D-xylulose 5-phosphate + NADPH + H ⁺ / 2-C-methyl-D-erythritol-4-phosphate + NADP ⁺	Co ²⁺ ; Mg ²⁺ ; Mn ²⁺	Takahashi et al., 1998	EG12715
IspD/ <i>ispD</i>	2-C-methyl-D-erythritol-4-phosphate + CTP + H ⁺ / 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol + diphosphate	CTP; Co ²⁺ ; Mg ²⁺ ; Mn ²⁺	Rohdich et al., 1999	G7423
IspE/ <i>ispE</i>	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol + ATP / 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol + ADP + 2 H ⁺	ATP;	Lüttgen et al., 2000; Lange & Croteau 1999	EG11294
IspF/ <i>ispF</i>	2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol / CMP + 2-C-methyl-D-erythritol-2,4-cyclodiphosphate	Mn ²⁺	Herz et al. 2000	EG11816
IspG/ <i>ispG</i> (<i>gcpE</i>)	2-C-methyl-D-erythritol-2,4-cyclodiphosphate + [reduced ferredoxin] ₂ / 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate + [oxidized ferredoxin] ₂ +	Mn ²⁺ ; flavodoxin or ferredoxin; NAD(P)H; [4Fe-4S] cluster	Altincicek et al., 2001; Seemann et al., 2002; Zepeck et al.,	EG10370

	H ₂ O		2005
IspH/ <i>ispH</i> (<i>lytB</i>)	NADPH + 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate + H ⁺ / NAD(P) ⁺ + dimethylallyl diphosphate + isopentenyl diphosphate + H ₂ O	Co ²⁺ ; flavodoxin or ferredoxin; NAD(P)H; [4Fe-4S] cluster	Rohdich et al., EG11081 2002; Seemann and Rohmer, 2007

ISC proteins are ubiquitously found throughout nature, but the approach of assembling these varies from species to species. Three different ISC-assembly pathways have so far been discovered in bacteria, namely the NIF (nitrogen fixation), ISC (iron-sulfur cluster) and SUF (sulfur mobilization) [Tokumoto et al. 2004], whereas a conserved ISC assembly machinery located in the mitochondria is present in eukaryotes. This machinery has been inherited from a bacterial ancestor of the mitochondria and therefore has high similarity to the ISC-assembly machinery found in bacteria [Mühlenhoff and Lill, 2000]. However, located in the mitochondria, the eukaryotic ISC-assembly machinery has to supply ISCs to apo-proteins throughout the cell. This is done by the cytosolic iron-sulfur cluster assembly machinery (CIA) [Lill and Mühlenhoff, 2008]. The CIA machinery is not very well characterized, and the exact mechanisms responsible for targeting cytosolic and nuclear apo-proteins to be loaded with ISCs exported from mitochondria are unknown. This obviously complicates engineering of the *ispG* and *ispH* to functionalize them within the cytosol of *S. cerevisiae*; in order to remedy this situation we decided to co-express the *E. coli* ISC-assembly machinery with the heterologous MEP-pathway.

The core functions in ISC biogenesis in *E. coli* are encoded by a single gene cluster composed of the following genes: *iscR* – transcriptional regulator, *iscS* – cysteine desulfurase (sulfur donor), *iscU* – scaffold, *iscA* – A-type protein, *HscB* – DnaJ-like co-chaperone, *HscD* – DnaK-like chaperone, *fdx* – ferredoxin and *iscX* – unknown function, found to interact with *iscS* [Py and Barras, 2010; Tokumoto et al., 2002]. Besides these 10 genes, the *ErpA* gene, encoding another A-type protein, is essential in *E. coli* for converting *ispG* from its apo form to its holo form [Loiseau et al., 2007]. Furthermore, the *E. coli* CyaY gene, encoding frataxin, takes part in ISC-assembly, probably as an iron-donor [Yoon and Cowan, 2003]. Bedekoviks et al. [2007] showed that a mitochondrially targeted version of the *E. coli* CyaY gene could functionally complement the *yfh1* deletion in *S. cerevisiae*. *YFH1* has previously been found to be important for ISC-biogenesis in yeast, hereby further strengthening the evidence that frataxin is required for ISC-assembly [Duby et al., 2002]. Finally, Justino et al. [2007] showed that the *E. coli* YtfE-encoded protein acts as a repair protein to mend ISC-proteins that have been damaged due to nitrosative stress.

Iron storage and sequestration in biological system is of outmost importance to ensure that the iron is soluble and thereby accessible where necessitated and to avoid the formation of reactive oxygen species, e.g. through the Haber-Weiss-Fenton reaction, which generates hydroxyl radicals [Aisen et al., 2001]. Most organisms use the polymeric protein ferritin to store iron. Even though *S. cerevisiae* contains ferritin, the majority of the iron stored within the cell is located in the vacuole [Raguzzi et al., 1988]. Little is known about how iron is transported within the yeast cell, and it is unknown whether the iron stored in the vacuole will be available for a heterologous ISC-machinery. It has previously been found that iron storage in *S. cerevisiae*, and thereby tolerance towards stress induced by high extracellular iron concentrations, can be improved by heterologous expression of human hetero-polymeric ferritin [Kim et al., 2003]. We therefore decided to additionally express the *E. coli* *Bfr*, *FtnA*, and *Bfd* genes, encoding the bacterioferritin, ferritin and bacterioferritin-associated ferredoxin, respectively. These iron storage genes together with the ISC-machinery genes, the ISC repair gene and the *Fdx*, *Fpr*, and *FldA*, encoding the ferredoxin, ferredoxin (flavodoxin) NADP⁺ reductase, and flavodoxin 1 [Nakamura et al., 1999; Bianchi et al., 1993; Puan et al., 2005], make up the supporting system that we hypothesized to be required to functionalize the MEP-pathway expressed in *S. cerevisiae*. Here we report the cloning and expression of these 16 genes combined with the 8 MEP-pathway genes (including *Idi*), together with the testing of this complex system. We finally show that *S. cerevisiae* is capable of functional expression of bacterial ISC-proteins within the cytosol.

2.3 Materials and Methods

2.3.1 Chemicals, enzymes and growth media

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich®, St. Louis, MO, and all enzymes from New England Biolabs Inc., Ipswich, MA.

E. coli cells were grown in BD Difco™ LB broth from BD Diagnostic Systems, Sparks, MD. For selection and maintenance of plasmids, ampicillin was added to a final concentration of 100 mg/l. *S. cerevisiae* cultures that required no auxotrophic selection were grown on YPD, composed of: 10 g/l Bacto™ yeast extract, BD Diagnostic Systems; 20 g/l Bacto™ peptone, BD Diagnostic Systems; and 20 g/l dextrose. For auxotrophic selection, *S. cerevisiae* was grown on SC-medium with the appropriate nutrient(s) omitted or SD for prototrophic strains. SC medium was composed of 6.7 g/l Yeast Nitrogen Base (YNB) without amino acids, BD Diagnostic Systems; CSM – amino acid dropout mixture (amount according to manufacturer's recommendation), Sunrise Science Products, Inc., San Diego, CA; and 20 g/l dextrose or galactose. SD was made the same way as SC with the difference of omitting the CSM amino acid dropout mixture. All media components and media besides YNB were sterilized by autoclaving. YNB was sterilized by filtration.

2.3.2 Plasmids and strains

The strains and plasmids acquired for this study can be seen from Table 2.2. Every cloning for the construction of the plasmids described below was carried out by treating the recipient plasmids with endonuclease(s), purification by gel-electrophoresis and recovery by gel extraction using the PureLink™ Quick Gel Extraction Kit, Life Technologies, Grand Island, NY, followed by ligation to the insert using T4 DNA ligase. Each insert was either prepared by removing it from the plasmid harboring it by endonuclease digestion followed by gel-electrophoresis and purification in the same way as the plasmid recipient, or amplified by PCR, purified using the QIAquick PCR Purification Kit, Valencia, CA, treated with endonucleases and purified again using the QIAquick PCR Purification Kit. All constructs were verified by sequencing. The DNA ladder used as size reference was the Quick-Load® 1 kb DNA Ladder from New England Biolabs Inc., Ipswich, MA (bands corresponds to the following DNA lengths: 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 8.0, 10.0 kb).

Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

Table 2.2: Strains and plasmids acquired for this study.

Strain	Genotype	Source
<i>Saccharomyces cerevisiae</i>		
F1702	<i>MATa ade1 arg4 his2 leu2 trp1 ura3</i>	Dr. G. Fink, Whitehead Institute, Cambridge, MA
CEN.PK 113-7D	<i>MATa MAL2-8^c SUC2</i>	Dr. P. Kötter, Frankfurt, Germany.
JAY20	<i>MATa leu1 ura3-52</i> (derived from FY2, Winston et al. 1995)	Dr. J. Avalos and Dr. G. Fink, Whitehead Institute, Cambridge, MA
<i>Escherichia coli</i>		
MAX Efficiency® DH5α™	F- φ80 <i>lacZΔM15</i> Δ(<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (rk-, mk+) <i>phoA supE44 λ- thi-1 gyrA96 relA1</i>	Life Technologies, Grand Island, NY
K-12, MG1655	F- lambda- <i>ilvG- rfb-50 rph-1</i>	Ajikumar et al., 2010
Plasmid name		Origin/reference
pKT127		Sheff and Thorn, 2004.
pRS414TEF		Mumberg et al., 1995
pRS415GPD		Mumberg et al., 1995
PUCAR1		Dr. Hang Zhou
pUC19		Life Technologies, Grand Island, NY
PUG72		Güldener et. al., 2002
pESC-URA		Stratagene, Agilent Technologies, Inc., Santa Clara, CA

2.3.3 Construction of expression cassettes

For expression of the genes of interest, four expression cassettes were created and tested for their ability to mediate transcription by verifying that they could express the yeast enhanced green fluorescent protein (yeGFP). This was done on the basis of the pRS414TEF and pRS415GPD in which the yeGFP from pKT127 was inserted by SpeI/XhoI digestion. The pRS414TEF with the insert yeGFP was modified by exchanging the CYC1t with the ADH1t, ACT1t and ENO2t by XhoI/KpnI digestion. The TEF1p was substituted with the PGK1p in the plasmid with the ACT1t, and with the TEF2p in the plasmid with the ENO2t, by SacI/SpeI digestion. The replacement promoters and terminators were obtained by PCR from CEN.PK 113-7D genomic DNA (gDNA) using the primers containing adaptamers with the respective restriction sites (see Table 2.3). This resulted in the construction of 4 expression cassettes harboring yeGFP, namely the *TEF1p-yeGFP-ADH1t*, *TDH3p (GPDp)-yeGFP-CYC1t*, *PGK1p-yeGFP-ACT1t* and *TEF2p-yeGFP-ENO2t* in the pSCX001, pSCX002, pSCX003 and pSCX004 plasmids, respectively.

Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

Table 2.3: Primers.

Feature	Forward ¹ primer (5' - 3')	Reverse ¹ Primer (5' - 3')	Amplicon size (bp) ²	Restriction sites ³
<i>yeGFP</i>	GACTTA <u>ACTAGT</u> ATGT CTAAAGGTGAAGAATT ATTCACTGG	TATGAC <u>CTCGAG</u> TT ATTTGTACAATTCAT CCATACCATGGG	717	SpeI/XhoI
<i>ADH1t</i>	GACTTA <u>CTCGAG</u> GCGA ATTTCTTATGATTTAT GATTTTTATTATTAAA TAAGT	TATGAC <u>GGTACCGA</u> GCGACCTCATGCTA TACCTG	166	XhoI/KpnI
<i>PGK1p</i>	GACTTAG <u>AGCTCAGAC</u> GCGAATTTTTCGAAGA AGTACC	TATGAC <u>ACTAGT</u> TG TTTTATATTTGTTGT AAAAAGTAGATAAT TACTTCC	1000	SacI/SpeI
<i>ACT1t</i>	GACTTA <u>CTCGAGTCTC</u> TGCTTTTGTGCGCGTA TGT	TATGAC <u>GGTACCAT</u> ATGATACACGGTCC AATGGATAAAC	293	XhoI/KpnI
<i>TEF2p</i>	GACTTAG <u>AGCTCGGGC</u> GCCATAACCAAGGTAT C	TATGAC <u>ACTAGT</u> GT TTAGTTAATTATAGT TCGTTGACCGTATA	500	SacI/SpeI
<i>ENO2t</i>	GACTTA <u>CTCGAGAGTG</u> CTTTAACTAAGAATT ATTAGTCTTTTCTG	TATGAC <u>GGTACCAG</u> GTATCATCTCCATCT CCCATATG	400	XhoI/KpnI
<i>HIS2</i>	GACTTAA <u>AGCTTTCAA</u> TCTTGCCGGTTTCATA CATGTTG	TATGACA <u>AGCTTAA</u> AGGTTTCCCAGCCA AACCCG	1750	HindIII
<i>ADE1</i>	GACTTAA <u>AGCTTATTC</u> ACGAGTCAGTCTGACT CTTG	TATGACA <u>AGCTTGA</u> ATACGAAAGAGAAC TGAGTCAGTTG	1427	HindIII
<i>YPRCδ15-UP</i>	GACTTAG <u>ACGTCTTAA</u> <u>TTAAGCCAGGCGCCTT</u> TATATCAT	TATGAC <u>GTCGACGC</u> <u>GGCCGCTGACCTAG</u> <u>ACCTAGGTTTGCGA</u>	690	AatII- PacI/SalI- NotI-AvrII

		AACCCTATGCTCT		
<i>YPRCδ15-DOWN</i>	GACTTAGCTAGCAATG GAAGGTCGGGATGAG	TATGACGCGGCCGC TGACCTAGACCTAG GATAAAGCAGCCGC TACCAA	651	NheI/NotI- AvrII
<i>YPRCτ3-UP</i>	GACTTAGACGTCTTAA TTAAAAAGGAGGTGC ACGCATTAT	TATGACGTCGACGC GGCCGCTGACCTAG ACCTAGGTTCCAAG GAGGTGAAGAACG	598	AatII- PacI/SalI- NotI-AvrII
<i>YPRCτ3-DOWN</i>	GACTTAGCTAGCGATG GGACGTCAGCACTGTA	TATGACGCGGCCGC TGACCTAGACCTAG GCGGTATTACTCGA GCCCCGTA	638	NheI/NotI- AvrII
Expression cassettes	GACTTAGCTAGCCCTC ACTAAAGGGAACAAA AGCTG	TATGACGCGGCCGC TGACCTAGACCTAG GAATACGACTCACT ATAGGGCGAATTG	Variable	NheI/NotI- AvrII
<i>TRP1</i>	GACTTAGCTAGCGTAC AATCTTGATCCGGAGC TTTTTC	TATGACGCGGCCGC TGACCTAGACCTAG GAGGCAAGTGCACA ACAATACTTAAAT AAATA	870	NheI/NotI- AvrII
<i>LEU2</i>	GACTTAGCTAGCTCGA CGGTCGAGGAGAACTT C	TATGACGCGGCCGC TGACCTAGACCTAG GTCGACTACGTCGT AAGGCCG	2235	NheI/NotI- AvrII
<i>ERG13-UP</i>	CGGAATAAGCTTCGTA TATACAATAGAAAAAT TTTC	TATGACCTGCAGGC TGCACCTTTTATAGT AATTTGGCTAC	213	HindIII/ PstI
<i>ERG13-DOWN</i>	GACTTA ACTAGT CCCG GGTCTTCCCCATCGA TTGCATCTTG	TATGACCCGCGGCG GCCGGGAAAACACG TCGGGGTTATGAAT G	272	SpeI/SacII
<i>ERG13-</i>	CACTATCCATCCGACA	ACCAAGGGGATTAT	1000	

Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

Delta(1)	GATGGAC	TGATGCTTGCT		
<i>ERG13</i> - Delta(2)	TTGAGATGAGCTTAAT CATGTCAAAGC	GATGTTTCTAAAGT GGCCTGTACG	900	
<i>ERG13</i>	ATGAAACTCTCAACTA AACTTTGTTGGTG	TTTTTTAACATCGTA AGATCTTCTAAATTT GTCATC	1473	
<i>leuC</i>	GACTTAGGATCCATGG CTAAGACGTTATACGA AAAATTGTTC	TATGACAAGCTTTT ATTTAATGTTGCGA ATGTCGGCGAAATG	1401	BamHI/ HindIII
<i>leuD</i>	GACTTAGAATTCATGG CAGAGAAATTTATCAA ACACACAG	TATGACGAGCTCTT AATTCATAAACGCA GGTTGTTTTGCTTC	606	EcoRI/SacI
<i>LEU1</i>	GACTTAGGATCCATGG TTTACACTCCATCCAA GGG	TATGACCTCGAGCT ACCAATCCTGGTGG ACTTTATC	2340	BamHI/ XhoI
<i>TEF1p-dxs- ADH1t</i>	CGAGGGGCCCCTGCTGG AGCTCATAGCTTC	CCAGCGACGTCTAC CGAGCGACCTC	2483	ApaI/AatII
<i>PGK1p- ispF-ACT1t</i>	CCCAGCATGCAGACGC GAATTTTTCG	CTAGACCTAGGGCG AATTGGGTACCGGC CGC	1806	SphI/AflIII
<i>TDH3p- ispE-CYC1t</i>	CGGTAGACGTCGCTGG AGCTCAGTTTATC	CTAGACCTAGGGCG AATTGGGTACCGG	1823	AatII/AvrII
<i>TEF1p- ispD- CYC1t</i>	GCTAGGCGCCGGGAA CAAAAGCTGGAGCTC ATAGC	GTACCATATGCGAC TCACTATAGGGCGA ATTGGG	1450	KasI/NdeI
<i>TEF1p- ispG- CYC1t</i>	CTGCAGGCATGCAAGC TTGAATACGAAAGAG	GACCGTGTATCATA TCGGCCGGTACCCA CTTAAGGGC	1879	SphI/AflIII
<i>TDH3p- ispH- CYC1t</i>	GGTAGACGTCGCTGGA GCTCAGTTTATCATTA TC	GCTCCTAGGGCGAA TTGGGTACCGGCCG C	1919	AatII/AvrII

<i>PGK1p- ispC-ACT1t</i>	CGCCCTAGGAGCTGGA GCTCAGACGCGAATTT TTC	GCCCTTAAGTGGGT ACCGGCCGATATGA TACAC	2544	AvrII/AflII
<i>TEF1p-idi- ADH1t</i>	GAGGGGCCCCGCTGGA GCTCATAGCTTC	CAGCGACGTCTACC GAGCGACCTCATGC	1167	ApaI/AatII
<i>TDH3p- YtfE-CYC1t</i>	CTAGAGGATCCCCGGG TACCGCTGGAGCTCAG TTTATCATTATC	CACTTAAGGGCCGC AAATTAAAGCCTTC GAGCGTCCC	1626	BamHI/ AflII
<i>TEF1p-bfr- ADH1t</i>	CGAGGGGCCCCGCTGG AGCTCATAGCTTC	CCAGCGACGTCTAC CGAGCGACCTCATG CTATAC	1511	ApaI/AatII
<i>PGK1p- bfd-ACT1t</i>	CGCCCTAGGAGCTGGA GCTCAGACGCGA	GGCCCTTAAGTGGG TACCGGCCGATATG ATACACGGTCCAAT G	1546	AvrII/AflII
<i>TDH3p- FtnA- CYC1t</i>	CGGTAGACGTCGCTGG AGCTCAGTTTATCATT ATC	GCTCCTAGGGCGAA TTGGGTACCCGGCC GCAAATTAAAGCCT TCG	1468	AatII/AvrII
<i>ARG4</i>	CGCCAAGCTTCACTGT CAGAGACTGTTTCC	GAGGATCCAGAGTC GACAACGACTTTGG GAGG	2320	HindIII/ BamHI

¹ Primer direction defined according to the direction of translation of the ORF.

² Amplicon size does not include adaptamers containing the restriction sites.

³ Restriction site(s) in the adaptamer of the forward primer / restriction site(s) in the adaptamer of the reverse primer.

The 8 MEP-pathway genes were purchased codon optimized for *S. cerevisiae* from DNA 2.0, Menlo Park, CA, flanked by SpeI (5') and XhoI (3') restriction sites. Using these restriction sites the genes were cloned into expression cassettes (see Table 2.4). The 16 ISC-assembly machinery genes were amplified from *E. coli* K12, MG1655 gDNA using Phusion® High-Fidelity DNA Polymerase (applied for all PCR reactions described in section 2.2.) and primers listed in Table 2.4. The forward and reverse primers were designed with adaptamers containing SpeI and XhoI

restriction sites respectively. These restriction sites were used to clone the genes into their designated expression cassettes.

Table 2.4: ORFs and primers.

ORF name	Forward ¹ primer (5' - 3')	Reverse ¹ Primer (5' - 3')	EcoCyc accession number	Expression cassette
<i>dxs</i> ²	ATGTCATTTGATATTG CCAAGTACCC	CTATGCCAACCAAGCC TTAATTTTC	G6237	TEF1p/ADH1t
<i>dxr</i> ²	ATGAAACAATTAACG ATTCTTGGCTCAAC	TCAGGATGCTAACCTC ATTACTTCTTT	EG12715	PGK1p/ACT1t
<i>ispD</i> ²	ATGGCAACAACGCACT TGGATG	TTACGTGTTTTCTTGGT GAATTGTTCTTG	G7423	TEF1p/ADH1t
<i>ispE</i> ²	ATGAGAACTCAGTGGC CTAGTC	TTAAAGCATGGCTCTAT GAAGTGGG	EG11294	TDH3p/CYC1t
<i>ispF</i> ²	ATGAGAATTGGTCACG GCTTCG	CTATTTTCGTTGCCTTGA TCAACAAG	EG11816	PGK1p/ACT1t
<i>ispG</i> ²	ATGCATAACCAAGCAC CCATCC	TCATTTTTCAACTTGCT GTACATCTATCC	EG10370	TEF1p/ADH1t
<i>ispH</i> ²	ATGCAAATATTGTTGG CTAATCCCAG	TCAGTCGACTTCTCTAA TATCAACAC	EG11081	TDH3p/CYC1t
<i>idi</i> ²	ATGCAAAGTGGAGCAC GTAATTCTG	CTACTTTAACTGGGTGA ATGCAGAT	G7508	TEF1p/ADH1t
<i>iscX</i>	GACTTAACTAGTATGG GACTTAAGTGGACCGA TAG	TATGACCTCGAGTTATT CGGCCTCGTCCAGCC	EG12311	TEF2p/ENO2t
<i>iscR</i>	GACTTAACTAGTATGA GACTGACATCTAAAGG GCG	TATGACCTCGAGTTATT AAGCGCGTAACTTAAC GTCGATC	G7326	TEF1p/ADH1t
<i>iscS</i>	GACTTAACTAGTATGA AATTACCGATTTATCT CGACTACTC	TATGACCTCGAGTTAAT GATGAGCCCATTCGAT GCTG	G7325	PGK1p/ACT1t

Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

<i>iscA</i>	GACTTAACTAGTATGT CGATTACACTGAGCGA CAGT	TATGACCTCGAGTCAA ACGTGGAAGCTTTCGC CG	EG12132	PGK1p/ACT1t
<i>iscU</i>	GACTTAACTAGTATGG CTTACAGCGAAAAAGT TATCGAC	TATGACCTCGAGTTATT TTGCTTCACGTTTGCTT TTATAGTCC	G7324	PGK1p/ACT1t
<i>HscA</i>	GACTTAACTAGTATGG CCTTATTACAAATTAG TGAACCTG	TATGACCTCGAGTTAA ACCTCGTCCACGGAAT GGC	EG12130	TEF1p/ADH1t
<i>HscB</i>	GACTTAACTAGTATGG ATTACTTCACCTCTTT GGC	TATGACCTCGAGTTAA AAATCGAGCAGTTTTTC TTCGAGTTG	EG12131	TEF1p/ADH1t
<i>CyaY</i>	GACTTAACTAGTATGA ACGACAGTGAATTTCA TCGCC	TATGACCTCGAGTTAG CGGAAACTGACTGTTT CACC	EG11653	TEF2p/ENO2t
<i>ErpA</i>	GACTTAACTAGTATGA GTGATGACGTAGCACT GC	TATGACCTCGAGTTAG ATACTAAAGGAAGAAC CGCAAC	EG12332	TEF2p/ENO2t
<i>bfr</i>	GACTTAACTAGTATGA AAGGTGATACTAAAGT TATAAATTATCTCAAC	TATGACCTCGAGTCAA CCTTCTTCGCGGATCTG TG	EG10113	TEF1p/ADH1t
<i>FtnA</i>	GACTTAACTAGTATGC TGAAACCAGAAATGA TTGAAAACTTAAT	TATGACCTCGAGTTAGT TTTGTGTGTCGAGGGTA GAG	EG10921	TDH3p/CYC1t
<i>bfd</i>	GACTTAACTAGTATGT ACGTTTGTCTTTGTAA TGGTATCAG	TATGACCTCGAGTTATG CGGACTCCTTAAACTCC G	EG11181	PGK1p/ACT1t
<i>fdx</i>	GACTTAACTAGTATGC CAAAGATTGTTATTTT GCCTCATC	TATGACCTCGAGTTAAT GCTCACGCGCATGGTT GAT	EG11328	TDH3p/CYC1t
<i>fpr</i>	GACTTAACTAGTATGG CTGATTGGGTAACAGG	TATGACCTCGAGTTACC AGTAATGCTCCGCTGTC	EG10628	TDH3p/CYC1t

	CAAA	ATA		
<i>FldA</i>	GACTTAACTAGTATGG CTATCACTGGCATCTT TTTCG	TATGACCTCGAGTCAG GCATTGAGAATTTCGTC GAGA	EG10318	TDH3p/CYC1t
<i>YtfE</i>	GACTTAACTAGTATGG CTTATCGCGACCAACC TTTA	TATGACCTCGAGTCACT CACCCGCCAGCGCGC	G7866	TDH3p/CYC1t

¹ Primer direction defined according to the direction of translation of the ORF.

² These genes were purchased codon optimized with flanking restriction sites and sub-cloned so no primers were required to clone these genes into their individual expression cassettes. The primers listed here are therefore the primers used for verification of genomic inserts and RT-PCR.

2.3.4 Construction of plasmids for integration of multiple expression cassettes

To integrate the 8 MEP-pathway genes in the genome of *S. cerevisiae*, two plasmids with 4 expression cassettes each were constructed based on the pUC19 plasmid. A 1750 bp fragment containing the *HIS2* ORF with 352 bp upstream of the start codon and 390 bp downstream of the stop codon, and a 1427 bp fragment containing the *ADE1* ORF with 332 bp upstream of the start codon and 274 bp downstream of the stop codon, were amplified from CEN.PK113-7D derived gDNA by PCR with primers containing HindIII adaptamers (see Table 2.3) and separately cloned into pUC19 digested with HindIII. The direction of the insert was determined by sequencing. The AatII sites in the backbone of the constructed vectors and in the *HIS2* ORF were removed by introducing silent point mutations, using the QuikChange® Multi Site-Directed Mutagenesis Kit, Stratagene, La Jolla, CA. A new multiple cloning site (MCS) comprised of two complementary primers (5'-GCGCGAGCTCCTCGAGGGGCCCAAGCTTGACGTCGTCGACCCTAGGCTGCAGCTTAAGTCTAGAGGTACCGCGC-3', which were annealed and subsequently digested with SacI and KpnI, were inserted in the plasmids treated with SacI and KpnI to create pSC600 (with the *HIS2* marker) and pSC700 (with the *ADE1* marker).

dxs, *ispF*, *ispE*, and *ispD* were treated with SpeI and XhoI restriction endonucleases and inserted in the pSCX001, pSCX003, pSCX002 and pRS414TEF, respectively, hereby creating the following expression cassettes: *TEF1p-dxs-ADH1t*, *PGK1p-ispF-ACT1t*, *TDH3p-ispE-CYC1t*, and *TEF1p-ispD-CYC1t*. *ispG*, *ispH*, *ispC*, and *idi* were treated with SpeI and XhoI restriction endonucleases and inserted in the pRS414TEF, pSCX002, pSCX003, and pSCX001 respectively, hereby creating the following expression cassettes: *TEF1p-ispG-CYC1t*, *TDH3p-ispH-CYC1t*, *PGK1p-ispC-ACT1t*, and *TEF1p-idi-ADH1t*. The cassettes were amplified by PCR with the primers listed in Table 2.3, purified, and digested with the sets of restriction endonucleases also

stated in Table 2.3. The *dxs*, *ispF*, *ispE*, and *ispD* expression cassettes were in the order listed here sequentially inserted in pSC600 to create pSC604. The *ispG*, *ispH*, *ispC*, and *idi* expression cassettes were in the order listed here sequentially inserted in pSC700 to create pSC704.

For chromosomal integration of the expression cassettes containing the *bfr*, *FtnA*, *bfd* and *YtfE* genes, a plasmid based on the pUC19 derived pSC700 plasmid was applied. The genes were amplified from *E. coli* K-12, MG1655 gDNA by PCR with primers listed in Table 2.4, digested with SpeI and XhoI and purified. *YtfE* and *FtnA* were inserted into the pSCX002 plasmid treated with SpeI and XhoI by ligation; and *bfr* and *bfd* in the same way into pSCX001 and pSCX002, respectively. The following expression cassettes were hereby constructed: *TDH3p-YtfE-CYC1t*, *TEF1p-bfr-ADH1t*, *PGK1p-bfd-ACT1t*, and *TDH3p-FtnA-CYC1t*. The expression cassettes were amplified by PCR using the primers listed in Table 2.3, and digested with the sets of restriction endonucleases specified in Table 2.3. The *YtfE*, *bfr*, *bfd*, and *FtnA* expression cassettes were in the order listed here sequentially inserted into pSC600, whereafter the *ARG4* marker was inserted between the HindIII and BamHI sites to construct pSC804. The *ARG4* marker was prior to the insertion amplified from CEN.PK113-7D derived gDNA with primers designed to amplify a DNA sequence from 546 bp upstream of the start codon to 364 bp downstream of the stop codon (primers listed in Table 2.3), digested with HindIII and BamHI, and purified.

To ensure high expression of the 12 remaining ISC-assembly machinery genes, we created two plasmids with 6 of the expression cassettes each targeting them into the YPRC δ 15 and YPRC τ 3 loci respectively, since integration of two different *lacZ* expression cassettes into these Ty remnants has given high expression [Flagfeldt et al., 2009]. We first separately introduced DNA sequences homologous to the upstream parts of the YPRC δ 15 and YPRC τ 3 loci into pUC19. The DNA was amplified from F1702 derived gDNA using reverse primers containing adaptamers which introduced the AvrII and NotI restriction sites. These two sites were used to digest the two plasmids, and the two first expression cassettes were amplified with primers introducing a NheI restriction site in front of the promoter and an AvrII and a NotI restriction site following the terminator; since AvrII and NheI are compatible, this allowed the insert treated with NheI and NotI to be inserted into the two plasmids, leaving behind a new set of AvrII and NotI restriction sites. This approach was iteratively used (see Supplementary Figure 2.1) to insert the expression cassettes into the two plasmids, finally creating the plasmid pSCFeS107TRP, containing the *iscX*, *iscR*, *iscS*, *fdx*, *TRP1*, *cyaY*, and *iscU* expression cassettes; and pSCFeS207LEU, containing the *hscB*, *fpr*, *iscA*, *erpA*, *LEU2*, *hscA* and *fldA*. The last insert in each of the two plasmids was the DNA sequences homologous to the downstream of the YPRC δ 15 and YPRC τ 3 loci respectively. All primers used for the construction of pSCFeS107TRP and pSCFeS207LEU are listed in Table 2.3.

2.3.5 Construction of other plasmids

ERG13 was deleted by a cassette constructed by introducing a 213 bp sequence identical to the gDNA sequence upstream of the *ERG13* start codon in the HindIII and PstI restriction sites and a 272 bp sequence identical to the gDNA sequence downstream of the *ERG13* stop codon in the SpeI and SacI restriction sites of pUG72, hereby creating plasmid pSCΔ001. The sequences were amplified from F1702 derived gDNA using the primers listed in Table 2.3.

To test whether *leuC* and *leuD* can complement the *leu1* auxotrophy, three plasmids were constructed, namely pESC-leuC, pESC-leuC/D and pESC-LEU1. pESC-leuC was constructed from pESC-URA by inserting *leuC* between the BamHI and HindIII restriction sites. pESC-leuC-D was constructed by inserting *leuD* between the EcoRI and SacI of pESC-leuC. The pESC-LEU1 plasmid was constructed by inserting *LEU1* between the BamHI and XhoI restriction sites of pESC-URA. *LeuC* and *leuD* were amplified from *E. coli* gDNA and *LEU1* from F1702 gDNA using the primers listed in Table 2.3.

All final plasmids constructed in this study are listed in Table 2.5, and their maps can be seen from Figure 2.2 in the Supplementary material.

Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

Table 2.5: Plasmids constructed in this study.

Plasmid name	Origin	Auxotrophic marker	Description
pSCX001	pRS414TEF	<i>TRP1</i>	Plasmid for test of expression cassette with GFP
pSCX002	pRS414TEF	<i>TRP1</i>	Plasmid for test of expression cassette with GFP
pSCX003	pRS415GPD	<i>LEU2</i>	Plasmid for test of expression cassette with GFP
pSCX004	pRS414TEF	<i>TRP1</i>	Plasmid for test of expression cassette with GFP
pSC604	pUC19	<i>HIS2</i>	Plasmid for genomic integration of <i>dxs</i> , <i>ispD</i> , <i>ispE</i> and <i>ispF</i> expression cassettes
pSC704	pUC19	<i>ADE1</i>	Plasmid for genomic integration of <i>dxr</i> , <i>ispG</i> , <i>ispH</i> and <i>idi</i> expression cassettes
pSC804	pUCAR1	<i>ARG4</i>	Plasmid for genomic integration of <i>bfr</i> , <i>FtnA</i> , <i>bfd</i> and <i>YtfE</i> expression cassettes
pSCFeS107TRP	pUC19	<i>TRP1</i>	Plasmid for genomic integration of <i>iscX</i> , <i>iscR</i> , <i>iscS</i> , <i>fdx</i> , <i>CyaY</i> and <i>iscU</i> expression cassettes
pSCFeS207LEU	pUC19	<i>LEU2</i>	Plasmid for genomic integration of <i>HscB</i> , <i>fpr</i> , <i>iscA</i> , <i>ErpA</i> , <i>HscA</i> and <i>FldA</i> expression cassettes
pSCΔ001	pUG72	<i>URA3</i>	Plasmid for deletion of ERG13
pESC-leuC	pESC-URA	<i>URA3</i>	2μ plasmid for expression of <i>leuC</i>
pESC-leuC-D	pESC-leuC	<i>URA3</i>	2μ plasmid for expression of <i>leuC</i> and <i>leuD</i>
pESC-LEU1	pESC-LEU1	<i>URA3</i>	2μ plasmid for expression of <i>leu1</i>

2.3.6 Construction of yeast strains

All transformations of *S. cerevisiae* were carried out using the lithium acetate/polyethylene glycol method according to Gietz and Woods [2002]. Transformants were selected on SC with the appropriate nutrient(s) omitted. The constructed strains were verified by extraction of gDNA using the Wizard® Genomic DNA Purification Kit, Promega, Madison, WI, followed by PCR amplification of all the integrated ORFs. Self-replicating plasmids were used directly to transform *S. cerevisiae*, whereas the plasmid destined for genomic integration was linearized. Prior to the transformation of *S. cerevisiae* the plasmids pSC604, pSC704 and pSC804 were linearized by digestion with MscI, HpaI and AgeI endonucleases respectively to facilitate their integration into the non-functional markers present in the yeast genome. The pSCFeS107TRP and pSCFeS207LEU integration cassettes were removed from their respective plasmids by digestion with PacI and NotI endonucleases, and the *ERG13* deletion cassette was removed from the pSCΔ001 plasmid by NotI and SacII digestion. The order in which the strains were constructed by transformation with the individual plasmids or cassettes can be seen from Table 2.6.

Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

Table 2.6: *S. cerevisiae* strains constructed in this study

Strain	Parental strain	Auxotrophic markers	Integrated plasmid or cassette	Integration locus or <i>S. cerevisiae</i> origin of replication
SCX001	F1702	<i>ade1, arg4 his2 leu2 ura3</i>	pSCX001	CEN6-ARSH4
SCX002	F1702	<i>ade1 arg4 his2 leu2 ura3</i>	pSCX002	CEN6-ARSH4
SCX003	F1702	<i>ade1 arg4 his2 trp1 ura3</i>	pSCX003	CEN6-ARSH4
SCX004	F1702	<i>ade1 arg4 his2 leu2 ura3</i>	pSCX004	CEN6-ARSH4
SC704	F1702	<i>arg4 his2 leu2 trp1 ura3</i>	pSC704	YAR015W
SCMEP	SC704	<i>arg4 leu2 trp1 ura3</i>	pSC604	YFR025C
SCMEP-FeS1	SCMEP	<i>arg4 trp1 ura3</i>	pSCFeS207LEU	YPRC τ 3
SCMEP-FeS2	SCMEP-FeS1	<i>arg4 ura3</i>	pSCFeS107TRP	YPRC δ 15
SCMEP-FeS2-IS	SCMEP-FeS2	<i>ura3</i>	pSC804	YHR018C
SCMEP-FeS2-IS- Δ ERG13	SCMEP-FeS2-IS	Prototrophic	pSC Δ 001	YML126C
SC-pESC	JAY20	<i>leu1</i>	pESC-URA	2 μ
SC-leuC	JAY20	<i>leu1</i>	pESC-leuC	2 μ
SC-leuC-D	JAY20	Prototrophic	pESC-leuC-D	2 μ
SC-LEU1	JAY20	Prototrophic	pESC-LEU1	2 μ

2.3.7 Test of expression cassette functionality using *yeGFP*

The SCX001, SCX002 and SCX004 were grown on SC-TRP and SCX003 on SC-LEU plates at 30 °C until colonies had formed. To test the expression of *yeGFP* the colonies were visually inspected using the Safe Imager™ 2.0 Blue-Light Transilluminator, Life Technologies, Grand Island, NY, with corresponding filter.

2.3.8 Liquid chromatography – Mass spectrometry analysis of MEP-pathway metabolites

SCMEP and SCMEP-FeS2-IS were grown in 5 ml SC-ADE-HIS and SC-ADE-ARG-HIS-LEU-TRP respectively at 30 °C at 250 rpm orbital shaking until the OD₆₀₀ reached 1. Two ml of each culture were harvested by centrifugation followed by extraction and analysis according to [Zhou et al., in press].

2.3.9 Verification of pathway functionality using U-13C₆ glucose

Strain SCMEP-FeS2-IS-ΔERG13 was inoculated at OD₆₀₀ = 0.05 in 5 ml SD medium made with U-13C₆ glucose, Cambridge Isotope Laboratories, Andover, MA and grown aerobically for 2 days at 30 °C with 250 rpm orbital shaking. The cells were spun down at 3,000 rpm for 10 min, re-suspended in 2 ml 20% w/v sodium hydroxide in 50% ethanol, transferred to an 8 ml glass tube, and extracted according to Madsen et al. [2011]. The extract was dried, re-dissolved in 50 µl pyridine, derivatized by addition of 50 µl N,O-bis(trimethylsilyl) trifluoroacetamide (BSTFA) + 1 % trimethylchlorosilane (TMCS) according to Nielsen and Madsen [2000], and analyzed using GC-MS. A 1-µl sample was analyzed using splitless injection on a DB-35ms, 30 m, 0.25 mm, 0.25 µm column, Agilent Technologies, Santa Clara, CA, with helium as carrier gas. The temperature of the oven was initially held a 160 °C for 30 sec, then increased to 320 °C at 10 °C/min, and finally held at 320 °C for 3 min. Ergosterol and squalene standards were subjected to the same procedure for authentication.

2.3.10 Verification of gene expression by reverse transcriptase polymerase chain reaction (RT-PCR)

Five ml. *S. cerevisiae* cultures were grown at 30 °C until OD₆₀₀ = 1, where after the cells were harvested by centrifugation at 3000 g for 10 min and washed with TE-buffer. They were transferred to a 2-ml FastPrep® tube, MP Biomedicals LLC, Solon, OH, and re-suspended in 500 µl TE-buffer. Two hundred µl acid washed glass beads (425-600 µm), Sigma-Aldrich® and 500 µl acidic phenol-chloroform-isoamyl alcohol (125:24:1), pH 4.5, Life Technologies, Grand Island, NY, were added. The cells were disrupted using a Savant FastPrep® FP120 Cell

Disrupter, MP Biomedicals LLC. The phases were separated by centrifugation at 18,000 g for 10 min, and 450 ml of the aqueous phase were transferred to a micro-centrifuge tube. Total RNA was precipitated by addition of 45 μ l 3 M sodium acetate, pH 5.2 and 450 μ l isopropanol, rinsed twice with 500 μ l 80 % ethanol and re-dissolved in 90 μ l nuclease-free water, Qiagen, Valencia, CA. The RNA was treated with DNase I, extracted with 400 μ l acidic phenol-chloroform-isoamyl alcohol (125:24:1), pH 4.5 and 400 μ l TE-buffer pH 8.0, then precipitated by sodium acetate and isopropanol and finally re-dissolved in 100 μ l nuclease free water. This RNA served as template for RT-PCR using the IllustraTM, Ready-To-GoTM, RT-PCR Beads from GE Healthcare Biosciences, Pittsburgh, PA. The RNA solution was checked for DNA contamination by PCR using the Taq DNA polymerase. The PCR products were visualized by gel electrophoresis (0.75 % agarose in TAE) and staining by Sybr® Safe, Life Technologies, Grand Island, NY.

2.3.11 Evolution of *S. cerevisiae* by serial sub-culturing

Strain SCMEP-FeS2-IS- Δ ERG13 was inoculated from glycerol stock in 5 ml SC-ADE-ARG-HIS-LEU-TRP-URA supplemented with 10 g/l mevalonolactone. The strain was incubated at 30 °C for 5 days with 250 rpm orbital shaking. Hereafter the OD₆₀₀ was measured, and 5 ml SC-ADE-ARG-HIS-LEU-TRP-URA with 5 g/l mevalonolactone was inoculated at OD₆₀₀ = 0.1. This periodical re-inoculation was continued, reducing the mevalonolactone concentration with every re-inoculation. The mevalonolactone was prepared as a 500 g/l stock-solution according to Dimster-Denk et al. [1994] by dissolving it in dH₂O at 65 °C for 1 h followed by filter sterilization.

2.3.12 Microscopy

The morphology of cells from the serial sub-culturing was investigated using a Nikon Eclipse TE200-S microscope, Nikon Instruments Inc. Melville, NY. The bright field images were obtained at 1000x magnification.

2.3.13 Functional complementation of *leu1* auxotrophy

Strains SC-pESC, SC-leuC, SC-leuC-D and SC-LEU1 were streaked out on SC (galactose)-LEU-URA plates from glycerol stock and incubated at 30 °C for 4 days, and the growth was assessed and documented by photography of the plates.

2.4 Results and discussion

Based on the pRS-series of *S. cerevisiae*/*E. coli* shuttle plasmids constructed by Mumberg et al. [1995], we constructed 4 plasmids, each containing an expression cassette harboring the gene encoding yeGFP. *S. cerevisiae* F1702 was transformed with these plasmids, and their ability to mediate expression was verified by visual inspection of the resulting colonies, result not shown. The 4 expression cassettes then served as starting point for creating 24 individual expression cassettes harboring the genes listed in Table 2.4. After systematic assembly of the 24 expression cassettes into 5 plasmids designed for chromosomal integration, the plasmids were one by one integrated in the genome of F1702. The first two plasmids to be integrated were pSC604 and pSC704, harboring the codon optimized genes encoding the MEP-pathway enzymes to construct the strain designated SCMEP. gDNA from SCMEP was isolated and used as template for PCR reactions confirming that all the genes were indeed integrated, result not shown. To test whether the heterologous MEP-pathway could supply the IPP required to sustain growth, we tried to delete the *ERG13* gene using a deletion cassette. We were unable to obtain viable colonies from which the *ERG13* gene had been knocked out. Several false positive colonies were obtained, probably due to integration of the deletion cassette at a non-specific locus. The SCMEP strain was then subjected to analysis using LC-MS, by which we detected 3.2 ± 0.5 mg DXP per g dry cell weight (DCW) and 5.0 ± 3.1 mg MEC per g DCW (the number following the \pm is the standard deviation based on biological triplicates). These results prove that the MEP-pathway is active until MEC, but whether the last two steps are active is impossible to determine based on these results since no HMBPP (detection limit = 0.2 mg/l) could be detected and it is impossible to discriminate between the IPP originating from the MVA-pathway and the MEP-pathway. What can be concluded is that the MEP-pathway is insufficient at supplying precursors in quantities that can support growth. However, this is the first time that DXP and MEC have been produced and detected in a *S. cerevisiae* strain.

The last two steps of the MEP-pathway are catalyzed by ISC-proteins, which require an elaborate support machinery to load the apo-proteins with ISC and couple these to NAD(P)H cofactor metabolism for supplying reducing equivalents. To remedy this potential problem we delineated the proteins that we, based on the literature, hypothesized could be sufficient to decouple the ISC assembly system from *E. coli* and graft it into the cytosol of *S. cerevisiae*. The 16 genes encoding these proteins were by the plasmids pSCFeS107TRP, pSCFeS207LEU and pSC804 integrated in the genome of SCMEP to construct SCMEP-FeS2-IS. gDNA was isolated from SCMEP-FeS2-IS, and all 24 genes were verified by PCR to be inserted within its genome. The results of the PCR can be seen from Figure 2.2.

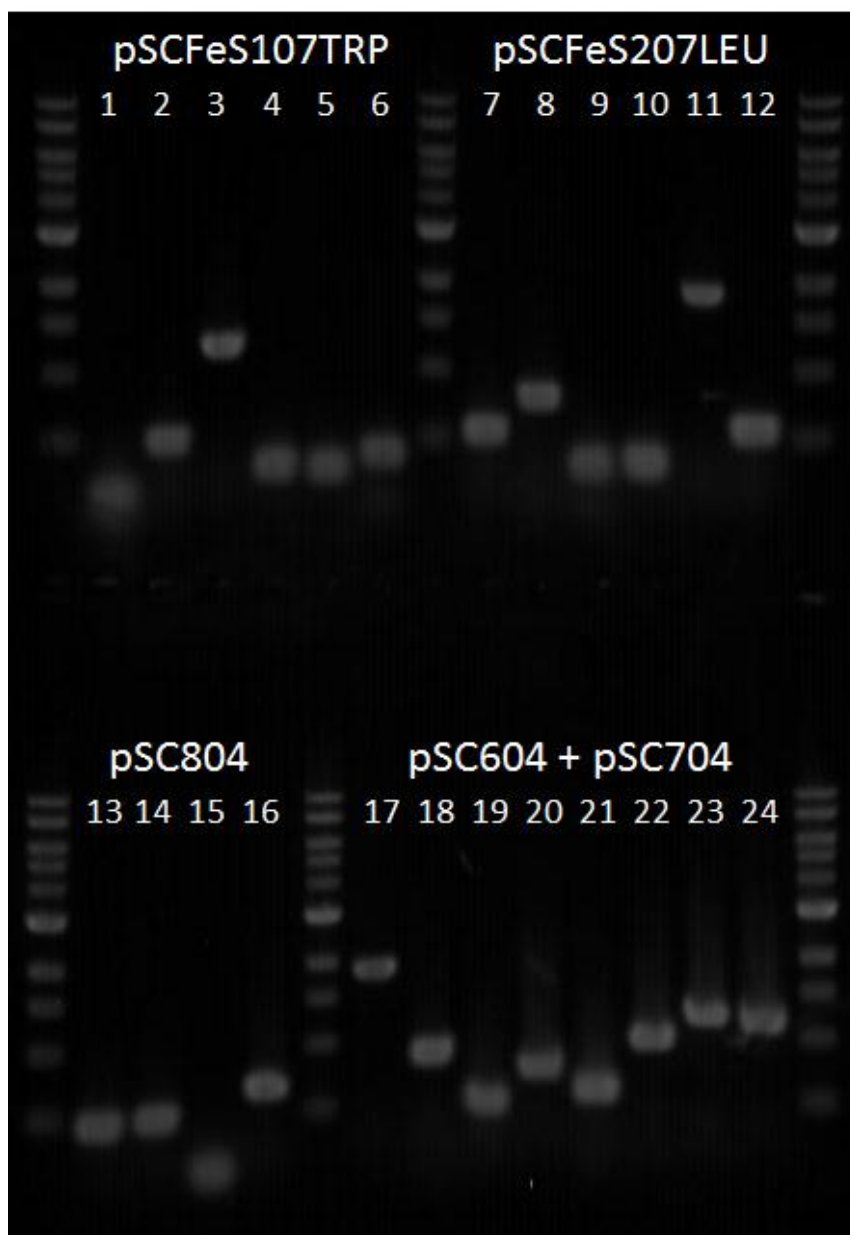


Figure 2.2: Verification of genomic integrated genes by PCR with SCMEP-FeS2-IS gDNA as template. To verify that all the genes had been inserted into the genome of SCMEP-FeS2-IS, gDNA was purified and used as template for amplification of the individual ORFs using primer pairs listed in Table 2.4. The amplification product was visualized by gel-electrophoresis on 0.75% agarose gel in TAE-buffer. Numbering of lanes corresponds to the following ORFs: 1: *iscX* (0.20 kb), 2: *iscR* (0.49 kb), 3: *iscS* (1.22 kb), 4: *fdx* (0.34 kb), 5: *CyaY* (0.32 kb), 6: *iscU* (0.39 kb), 7: *hscB* (0.52 kb), 8: *fpr* (0.75 kb), 9: *iscA* (0.32 kb), 10: *erpA* (0.35 kb), 11: *hscA* (1.85 kb), 12: *fldA* (0.53 kb), 13: *bfr* (0.48 kb), 14: *FtnA* (0.50 kb), 15: *bfd* (0.20 kb), 16: *YtfE* (0.66 kb), 17: *dxs* (1.86 kb), 18: *ispE* (0.85 kb), 19: *ispF* (0.48 kb), 20: *ispD* (0.71 kb), 21: *idi* (0.55 kb), 22: *ispH* (0.95 kb), 23: *dxr* (1.19 kb), 24: *ispG* (1.12 kb).

After verifying that all the genes had been inserted into the chromosome, we once more tried to delete the *ERG13* gene, but no viable colonies were obtained. This result does as aforementioned not prove that the MEP-pathway is inactive but only that the activity is too low to support growth, so we once again tested the strain using LC-MS, but we saw no significant change in MEC compared to the SCMEP strain. However the DXP concentration had increased to 7.9 ± 2.5 mg/gDCW; we do not know the reason for this increase. To be able to delete *ERG13* we had to supplement the medium with 10 g/L mevalonolactone, which serves as a source for mevalonate. Thirty colonies were screened by extraction of gDNA followed by PCR. Three colonies (colony 13, 25 and 29) were identified by PCR to have the deletion cassette integrated in the correct locus and no duplication of *ERG13* prior to the deletion, see Figure 2.3. These strains were named SCMEP-FeS-IS- Δ ERG13-strain13, SCMEP-FeS-IS- Δ ERG13-strain25 and SCMEP-FeS-IS- Δ ERG13-strain29.

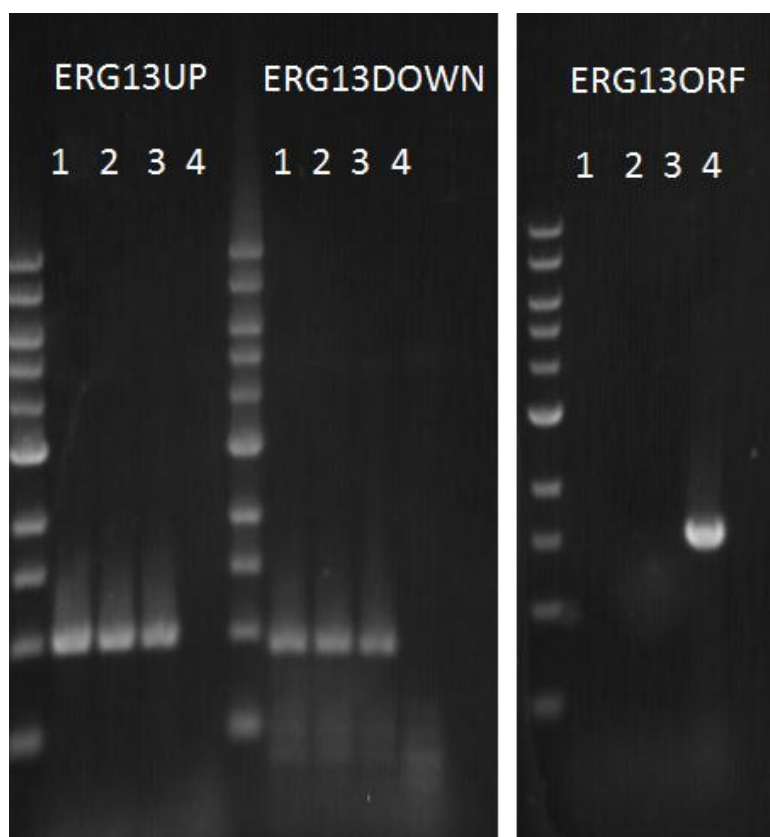


Figure 2.3: Verification of deletion of *ERG13*. gDNA purified from the following strains; 1: SCMEP-FeS2-IS- Δ ERG13-colony13, 2: SCMEP-FeS2-IS- Δ ERG13-colony25, 3: SCMEP-FeS2-IS- Δ ERG13-colony29 and 4: F1702 was used as templates for PCR reactions using the primers shown in Table 2.3. The PCR reactions of the ERG13UP (1.00 kb) and ERG13DOWN (0.90 kb) fragments verify that the deletion cassette has been inserted in the correct locus, and the inability to amplify the ERG13ORF (1.47 kb) from templates 1, 2 and 3 verifies that *ERG13* has not been duplicated prior to the knockout.

In order to find out if the MEP-pathway is active in the SCMEP-FeS2-IS-ΔERG13 strains isolated, they were grown on SD-medium made with uniformly C¹³-labelled glucose and unlabeled mevalonolactone as carbon sources. Since *ERG13* was deleted, the only way labeled carbon can end up in IPP is if the MEP-pathway is active. IPP is the first isoprenoid precursor of ergosterol, which is made from 2 FPP units (equivalent to 6 IPP units) forming squalene, which then undergoes several modifications to form ergosterol. During these modifications, 3 carbon atoms are lost due to the C14-demethylase and C4-decarboxylase activities encoded by *ERG11* and *ERG26*, respectively. Furthermore, a carbon is gained through the activity of the C24-methyl transferase, encoded by *ERG6* [Veen and Lang, 2004]. This means that if the MEP-pathway is active one would expect to see a shift in the molecular weight of ergosterol of 3 Daltons or higher, whereas a shift of 1 Dalton just indicates the labeled carbon gained from the C-24 methyl transferase. The ergosterol extracts from the 3 strains grown on medium with either uniformly C¹³-labelled glucose or regular glucose were analyzed on GC-MS and the peaks corresponding to squalene and ergosterol identified by comparisons of squalene and ergosterol standards. An example of the results obtained from this analysis can be seen from Supplementary Figure 2.3. The peak for the molecular weight of un-fragmented ergosterol derivatized with BSTFA is located at approximately 468 Daltons (corresponding to the M0 peak), and its eight mass-isotopomers peaks M1, M2, M3, M4, M5, M6, M7 and M8 (corresponding to M0 + increments of 1 Dalton respectively) were integrated for ergosterol isolated from the SCMEP-FeS2-IS-ΔERG13 strain grown on either uniformly C¹³-labelled glucose or regular glucose in triplicate. The results from all the strains were almost identical. The result from SCMEP-FeS-IS-ΔERG13-strain13 can be seen from Figure 2.4. The species that are heavier than M0 and M1 for the unlabeled and labeled, respectively, are due to natural occurrence of C¹³. These results show a shift of 1 Dalton from the unlabeled to the labeled, which indicates that the MEP-pathway has a very limited flux if any at all.

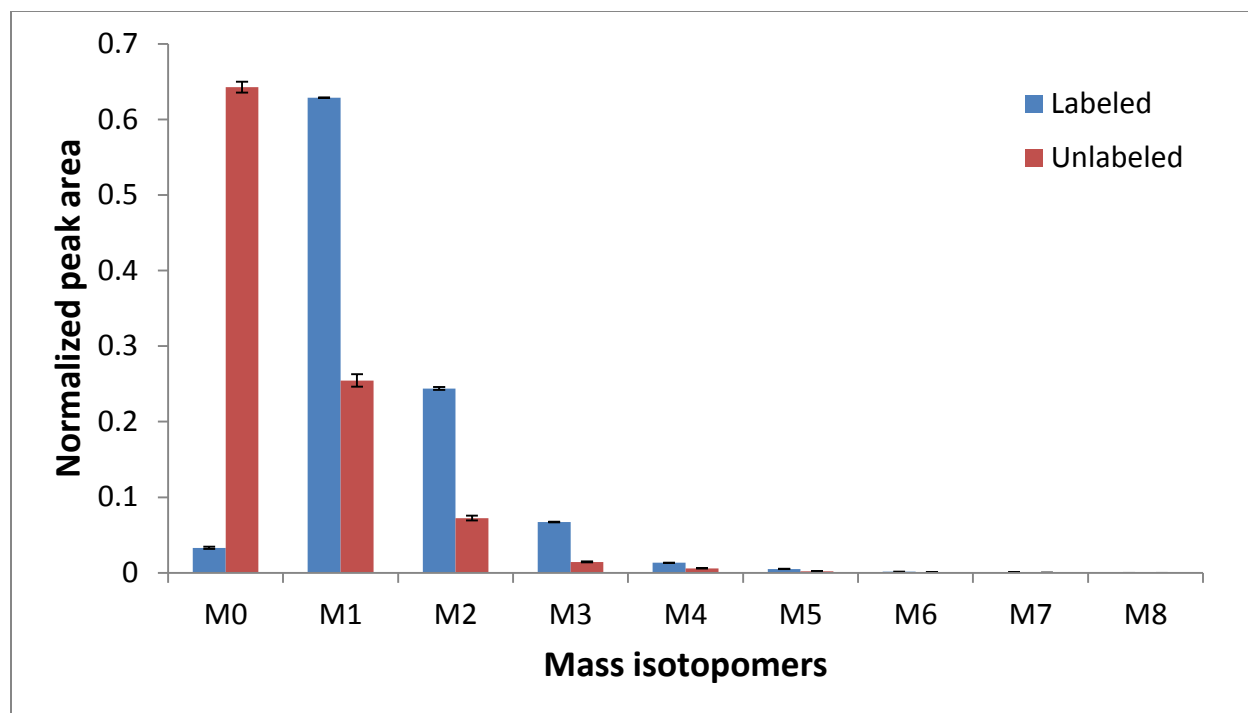


Figure 2.4: Ergosterol Mass isotopomer distribution. Integration of the peak areas of the ergosterol 468 Dalton (corresponding to M0) and 27 of its mass isotopomers (only the first 8 are shown since the area of every mass isotopomer peak above 5 is close to 0) at increments of 1 Dalton normalized to the total abundance. The labeled results are from ergosterol extracted from SCMEP-FeS-IS- Δ ERG13 grown on U- $^{13}\text{C}_6$ glucose whereas the unlabeled results are from SCMEP-FeS-IS- Δ ERG13 grown on regular glucose. The experiment was done in biological triplicates.

To test if all 24 genes were expressed, total RNA was isolated from SCMEP-FeS2-IS and tested for residual DNA by PCR. No DNA was found (results not shown). The RNA was used as template in RT-PCR reactions with primers pairs (Table 2.4) designed to amplify the entire ORF of the individual genes. The result of the RT-PCR reactions (Figure 2.5) showed that, except for *bfd*, *fldA* and *hscA*, all the genes are expressed. The experiment was therefore repeated, and *bfd* and *fldA* was also shown to be expressed (data not shown). However, it was not possible to detect any band for *hscA*. Since the same 4 promoters were used to express the genes, and as all other genes expressed by the same promoter as *hscA*, were expressed, the failure to detect transcription could be due to low mRNA stability, either *in vivo* or post cell breakage. However, lack of *hscA* expression is presumably not detrimental for functionality of the ISC-assembly machinery, since it has been found *in vitro* to be stimulatory rather than absolutely required for the transfer of ISC from the assembly machinery to the apo-proteins [Mansy et al., 2002].

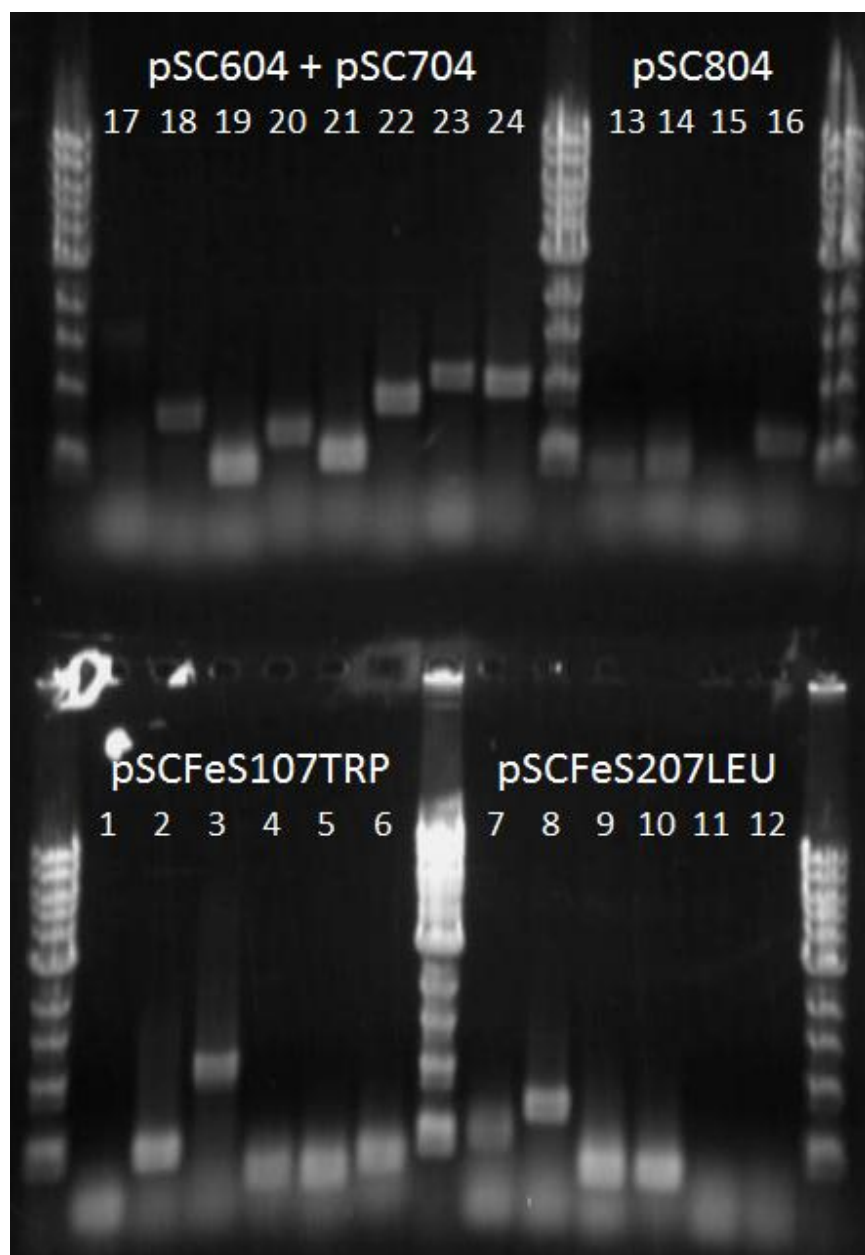


Figure 2.5: Verification of gene expression by Reverse Transcription PCR. To find out if the genes inserted in the genome of SCMEP-FeS2-IS are transcribed, total RNA was purified and used as template for amplification of the individual ORFs by RT-PCR using the primer pairs listed in Table 2.4. The amplification product was visualized by gel electrophoresis on 0.75% agarose gel in TAE-buffer. Numbering of lanes corresponds to the following ORFs: 1: *iscX* (0.20 kb), 2: *iscR* (0.49 kb), 3: *iscS* (1.22 kb), 4: *fdx* (0.34 kb), 5: *CyaY* (0.32 kb), 6: *iscU* (0.39 kb), 7: *hscB* (0.52 kb), 8: *fpr* (0.75 kb), 9: *iscA* (0.32 kb), 10: *erpA* (0.35 kb), 11: *hscA* (1.85 kb), 12: *fldA* (0.53 kb), 13: *bfr* (0.48 kb), 14: *FtnA* (0.50 kb), 15: *bfd* (0.20 kb), 16: *YtfE* (0.66 kb), 17: *dxs* (1.86 kb), 18: *ispE* (0.85 kb), 19: *ispF* (0.48 kb), 20: *ispD* (0.71 kb), 21: *idi* (0.55 kb), 22: *ispH* (0.95 kb), 23: *dxr* (1.19 kb), 24: *ispG* (1.12 kb).

To investigate the possibility that the SCMEP-FeS2-IS- Δ ERG13 strains have a finite, albeit low, flux through the MEP-pathway that could be increased, we set out to adapt the cultures to grow with IPP being solely supplied by the MEP-pathway by serial sub-culturing with decreasing amounts of mevalonolactone. Adaptive evolution has previously been found efficient for growth of engineered *S. cerevisiae* strains on xylose [Zhou et al., 2012]. The result of the evolution can be seen from Figure 2.6.

Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

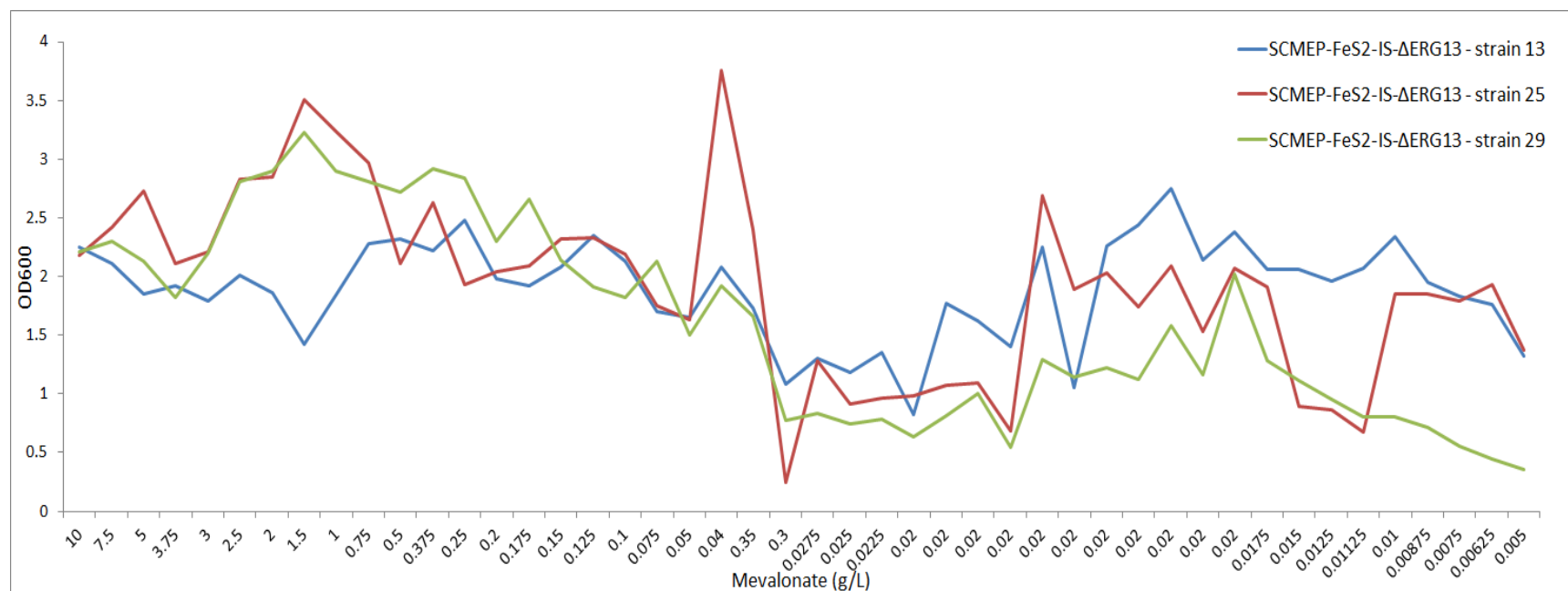


Figure 2.6: OD₆₀₀ measured during serial sub-culturing. Three strains of SCMEP-FeS-IS-ΔERG13 were grown on SC-ADE-ARG-HIS-LEU-TRP-URA supplemented with mevalonolactone. The cultures were inoculated at OD₆₀₀ = 0.1 at each sub-culturing and the mevalonolactone concentration gradually reduced. The OD₆₀₀ was measured after each sub-culturing which lasted for 5 days.

The newly created SCMEP-FeS2-IS- Δ ERG13 strains needed 5 g/l mevalonolactone to grow (results not shown), which corresponds with the value reported by [Umebayashi and Nakano, 2003]. However, serial sub-culturing for more than a half year resulted in two of the strains being able to grow to a similar OD₆₀₀ in medium with 5 mg/L mevalonolactone, i.e. 1000-fold less. Despite this strong adaptation no mevalonate independent cells could be isolated. In other words, the heterologous MEP-pathway was still unable to support growth. [Zhou et al., 2012] strain grew on xylose albeit slowly whereas the SCMEP-FeS2-IS- Δ ERG13 strains did not grow at all without mevalonolactone. This indicates that the MEP-pathway is probably inactive since it should be possible to evolve a pathway directly linked to growth with low activity to higher activity, whereas evolving a pathway with no activity to become active is less likely. During the directed evolution the morphology of the cells was observed using light microscopy. Figure 2.7 shows the morphology of the three SCMEP-FeS2-IS- Δ ERG13 strains grown in SC-ADE-ARG-HIS-LEU-TRP-URA + 0.25 g/l mevalonolactone. It can be seen that strain 25 and 29 exhibit abnormal morphology, probably due to the stress of low mevalonate availability.

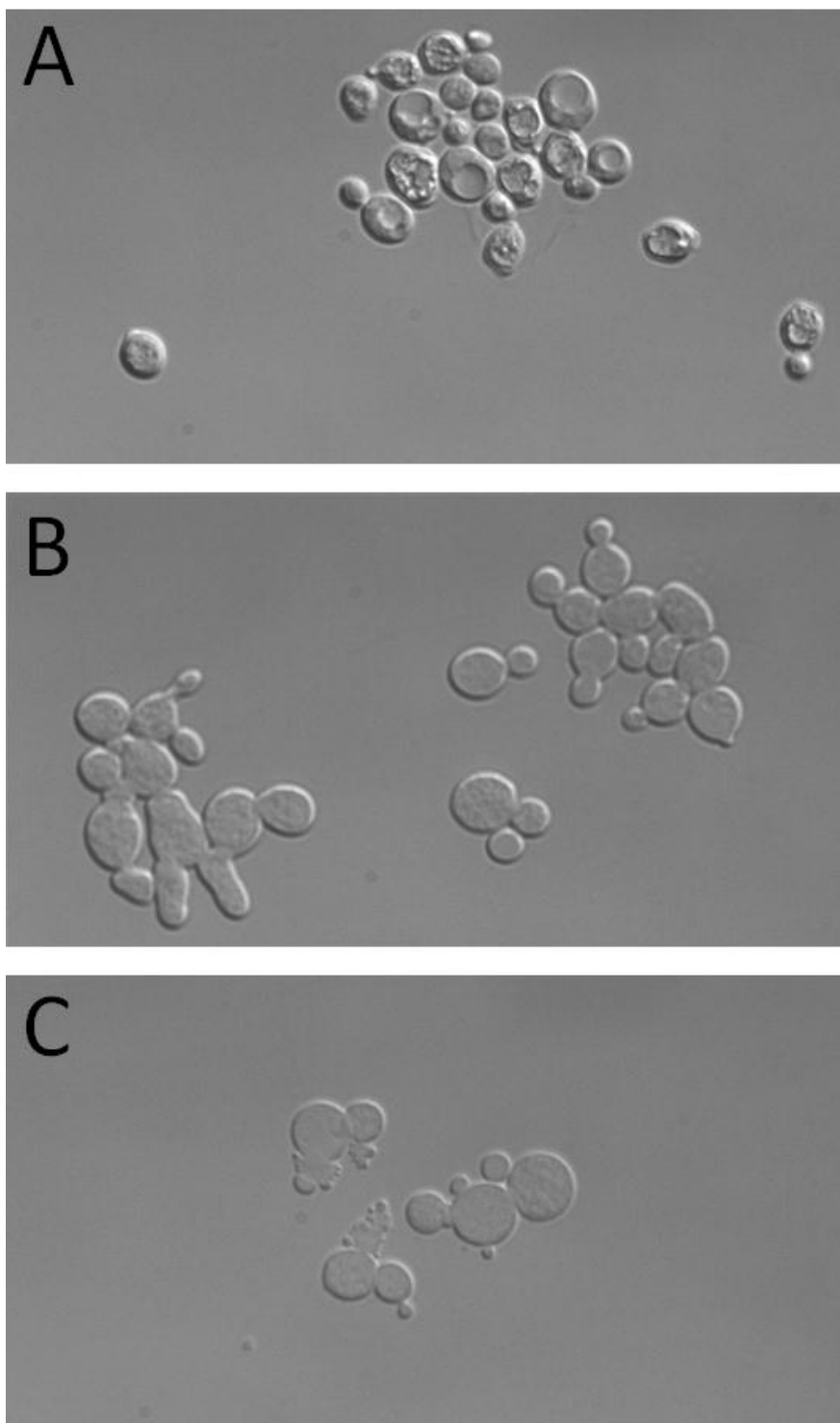


Figure 2.7: Microscope pictures of evolved SCMEP-FeS-IS- Δ ERG13 strains. The strains were grown in SC-ADE-ARG-HIS-LEU-TRP-URA + 0.25 g/l mevalonolactone for 5 days before the morphology was assessed by light-microscopy. A: SCMEP-FeS-IS- Δ ERG13-strain13, B: SCMEP-FeS-IS- Δ ERG13-strain25, and C: SCMEP-FeS-IS- Δ ERG13-strain29.

In order to test whether *S. cerevisiae* can functionally express other heterologous ISC-proteins from *E. coli* within its cytosol we decided to test if the *S. cerevisiae* strain JAY20 auxotrophic for leucine due to a deletion of the *LEU1* gene, encoding the second step in the biosynthesis of leucine, catalyzed by the isopropylmalate isomerase, could be complemented by the *E. coli* heterodimeric isopropylmalate isomerase, encoded by *leuC* and *leuD* [Skala et al., 1991 and Jardine et al., 2002]. The *leuC* gene was inserted in the pESC-URA plasmid followed by *leuD*, hereby constructing plasmids pESC-leuC and pESC-leuC-D, respectively. As a positive control, *LEU1* from F1702 was inserted in pESC-URA as well. These 3 plasmids together with the pESC-URA plasmid, which served as negative control, were used to transform JAY20. Expression of the genes integrated in pESC-URA was under control of the *GAL1/GAL10* inducible promoter system. The four strains were plated on SC-LEU-URA with galactose as carbon source. The result can be seen from Figure 2.8 and shows that pESC-leuC-D can complement the *LEU1* auxotrophy. This is to our knowledge the first example of a bacterial ISC-protein being functionally expressed in the cytosol of *S. cerevisiae*. *leuC* and *leuD* do however have a high degree of homology with *LEU1*, which could explain why they can be loaded with ISC. The mechanism of targeting apo-ISC-proteins to be loaded with ISC and hereby converted to their active holo-form has not yet been elucidated, but this result demonstrates that bacterial ISC-proteins heterologously expressed in *S. cerevisiae* can be converted to their functional holo-form.

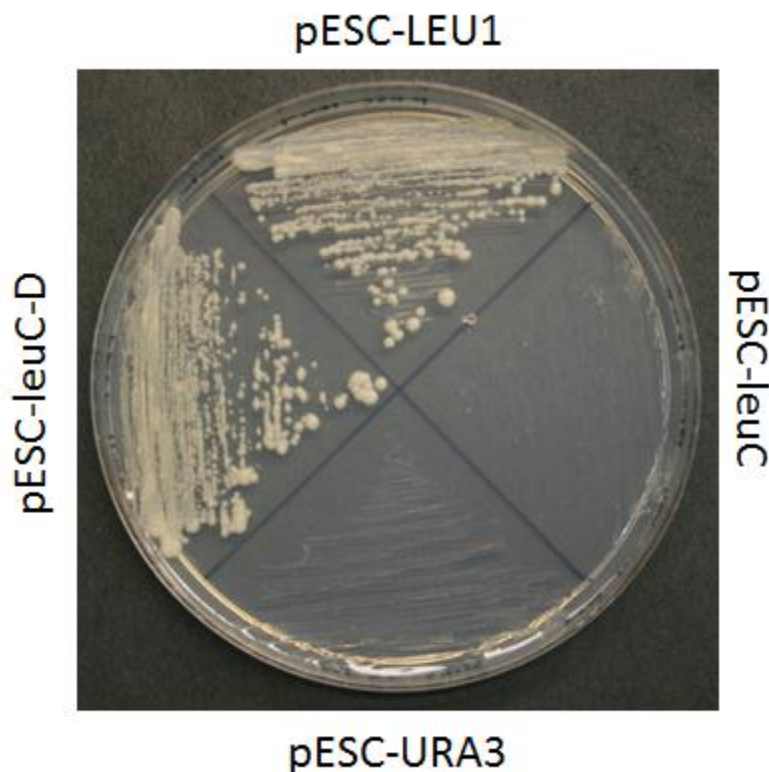


Figure 2.8: Complementation of the *leu1* auxotrophy in the JAY20 *S. cerevisiae* strain. JAY20 transformed with pESC-URA, pESC-leuC-D, pESC-LEU1 and pESC-leuC was plated on SC-LEU-URA and incubated at 30 °C for 3 days.

The fact that *S. cerevisiae* is capable of loading cytosolic heterodimeric isopropylmalate isomerase encoded by *leuC* and *leuD* with ISCs does not imply that all bacterial ISC-proteins heterologously expressed within yeast cytosol can be loaded with ISCs.

ispG and *ispH* are not the only two ISC-proteins among the proteins required for MEP-functionality. Ferredoxin is also an [2Fe-2S] ISC-protein [Ta and Vickery, 1992], which is involved in the transfer of electrons to *ispG* and *ispH*, as well as electron transfer to ISC-biogenesis. To be functional, ferredoxin therefore needs to be loaded with an ISC, which have to originate from the yeast CIA-machinery, since it cannot originate from the heterologous ISC-assembly machinery because no ISC can be generated through the heterologous expressed ISC-assembly machinery before ferredoxin is converted to its active holo-form. With ferredoxin as a pivotal point for MEP-pathway functionality, emphasis on research to assess whether this protein can be converted to its holo-form is required. If ferredoxin is found only in its apo-form, research should focus on adapting it to be loaded with ISC via the CIA-machinery, since this will be key to functionally express the MEP-pathway in the cytosol of *S. cerevisiae*.

The requirements for an elaborate ISC-assembly machinery for *ispG* and *ispH* function, combined with the finding that MEC can be detected in cell extract, suggest that the MEP-pathway is only active until MEC. This is further supported by the lack of labeled carbon ending up in ergosterol through the MEP pathway, as well as the mevalonolactone requirement of the *erg13Δ* strains. This view contradicts the view expressed by Maury et al. [2008], who used lovastatin in an attempt to find out if the MEP-pathway was active. However, the growth in the presence of lovastatin could have been caused by other circumstances, e.g., a mutation conferring resistance towards the drug. To circumvent the use of lovastatin, Maury et al. [2008] furthermore suggested deleting a gene encoding one of the early steps of the MVA-pathway such as *ERG10*, *ERG13* or *HMG1* and *HMG2*. We deleted *ERG13* in strain SCMEP, which only expresses the MEP-pathway genes, as well as in the SCMEP-FeS2-IS, expressing the MEP-pathway together with all the support genes hypothesized to be required. In neither case were we able to obtain transformants that had lost *ERG13* and could grow without mevalonolactone supplementation. Furthermore Maury et al. (2008) did not express any additional genes for coupling the *ispG* and *ispH* to NAD(P)H such as ferredoxin, flavodoxin and ferredoxin (flavodoxin) oxidoreductase. *S. cerevisiae* natively possesses ferredoxin, but it is located within the mitochondria and found to be required for the assembly of ISC-proteins [Lange et al., 2000]. No evidence of ferredoxin present in the cytosol of *S. cerevisiae* has been reported, and it is therefore probably not available for *ispG* and *ispH*. This is an additional reason to doubt that the MEP-pathway was functional in the *S. cerevisiae* system presented by Maury et al. [2008].

S. cerevisiae have several advantageous characteristics making it an ideal host for microbial production of isoprenoids such as well-established large scale fermentation, enabling easy scale-up, GRAS classification by the FDA, which eases the approval of novel products, compartmentalization for co-localization of enzymes, and strains pre-engineered to express cytochrome P450 which is important for the downstream decoration of the isoprenoid backbones have been demonstrated [Huang et al., 2008 and Truan et al., 1993]. The combination of *S. cerevisiae* with the stoichiometrical superior MEP-pathway holds great potential for bypassing native regulation, hereby constructing an excellent platform for isoprenoid production. Our results show that ISC-proteins can be expressed in the cytosol of yeast and that the MEP-pathway is active until MEC, which is encouraging for the further quest for functionalizing the MEP-pathway in its entirety. Further research into the requirements for targeting heterologous ISC-apo-proteins to be converted to their holo-form through the CIA machinery could help functionalize the MEP-pathway, whereas the approach of expressing the heterologous ISC-machinery in *S. cerevisiae* could not only help functionalize the MEP-pathway but also help us understand which parts of the *E. coli* ISC-machinery are essential and how this system should be delineated. Alone from *E. coli*, 100 ISC-proteins have been isolated [Fontecave, M. 2006], and being able to functionally express this group of ubiquitous proteins in *S. cerevisiae* encompasses great prospects for biosynthetic production of novel compounds, as well as obtaining a greater understanding of the complex processes involved in ISC-assembly. A *S. cerevisiae* strain pre-engineered to express the machinery required for cytosolic expression of heterologous ISC-

proteins would therefore be a great tool in many areas within biotechnological and pharmaceutical research. Our work has showed the initial steps towards obtaining such strains and highlighted the obstacles on the way towards reaching this goal. This will hopefully spark further interest within the area of heterologous expression of ISC-proteins.

2.5 Acknowledgements

The authors would like to thank Dr. Hang Zhou, Dr. Jose Avalos and Dr. Gerald R. Fink for contributing the plasmids and yeast strains used in this study. The authors furthermore gratefully acknowledge the funding from the Technical University of Denmark and Massachusetts Institute of Technology.

2.6 References

- Aisen, P.; Enns, C.; and Wessling-Resnick, M. Chemistry and biology of eukaryotic iron metabolism. *Int. J. Biochem. Cell B.* 33, 940-959 (2001).
- Ajikumar, P. K.; Tyo, K.; Carlsen, K.; Mucha, O.; Phon, T. P.; and Stephanopoulos, G. Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharm.* 5(2), 167-190 (2008).
- Ajikumar, P. K.; Xiao, W.; Tyo K. E. J.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; and Stephanopoulos, G. Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* 330:70-74 (2010).
- Alberts, A. W.; Chen, J.; Kuron, G.; Hunt, V.; Huff, J.; Hoffman, C.; Rothrock, J.; Lopez, M.; Joshua, H.; Harris, E.; Patchett, A.; Monaghan, R.; Currie, S.; Stapley, E.; Albers-Schonberg, G.; Hensens, O.; Hirshfield, J.; Hoogsteen, K.; Liesch, J.; and Springer, J. Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. *Proc. Natl. Acad. Sci.* 77, 3957-3961 (1980).
- Altincicek, B.; Kollas, A.; Sanderbrand, S.; Wiesner, J.; Hintz, M.; Beck, E.; and Jomaa, H. GcpE Is Involved in the 2-C-methyl-d-erythritol 4-phosphate pathway of isoprenoid biosynthesis in *Escherichia coli*. *J. Bacteriol.* 183(8), 2411-2416 (2001).
- Bedekoviks, T.; Gajdos, G. B.; Kisplá, G.; and Isaya, G. Partial conservation of functions between eukaryotic frataxin and the *Escherichia coli* frataxin homolog. *FEMS Yeast Res.* 7, 1276-1284 (2007).
- Bianchi, V.; Reichard, P.; Eliasson, R.; Pontis, E.; Krook, M.; Jörnvall, H.; and Haggård-Ljungquist, E. *Escherichia coli* ferredoxin NADP⁺ reductase: activation of *E. coli* anaerobic ribonucleotide reduction, cloning of the gene (*fpr*), and overexpression of the protein. *J. Bacteriol.* 175(6), 1590-1595 (1993).
- Blaschkowski, H. P.; Neuer, G.; Ludwig-Festl, M.; and Knappe, J. Routes of flavodoxin and ferredoxin reduction in *Escherichia coli*. CoA-acylating pyruvate: flavodoxin and NADPH: flavodoxin oxidoreductases participating in the activation of pyruvate formate-lyase. *Eur. J. Chem.* 123, 563-569 (1982).
- Dimster-Denk, D.; Thorsness, M. K.; and Rine, J. Feedback regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase in *Saccharomyces cerevisiae*. *Mol. Biol. Cell.* 5, 655-665 (1994).
- Duby, G.; Foury, F.; Ramazotti, A.; Herrmann, J.; and Lutz, T. A non-essential function for yeast frataxin in iron-sulfur cluster assembly. *Human Mol. Genet.* 11, 2635-2643 (2002).

Dugar, D.; and Stephanopoulos, G. Relative potential of biosynthetic pathways for biofuels and bio-based products. *Nat. Biotech.* 29, 1074–1078 (2011).

Flagfeldt, D. B.; Siewers, V.; Huang, L.; and Nielsen, J. Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast*. 26, 545–551 (2009).

Fontecave, M. Iron-sulfur clusters: ever-expanding roles. *Nat. Chem. Biol.* 2, 171-174 (2006).

Gietz, R. D.; and Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Method. Enzymol.* 350, 87–96 (2002).

Gueldener, U.; Heinisch, J. Koehler, G. J.; Voss, D.; and Hegemann, J. H. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucl. Acids Res.* 30(6), e23 (2002).

Herz, S.; Wungsintaweeikul, J.; Schuhr, C. A.; Hecht, S.; Lüttgen, H.; Sagner, S.; Fellermeier, M.; Eisenreich, W.; Zenk, M. H.; Bacher, A.; and Rohdich, F. Biosynthesis of terpenoids: YgbB protein converts 4-diphosphocytidyl-2C-methyl-d-erythritol 2-phosphate to 2C-methyl-d-erythritol 2,4-cyclodiphosphate. *Proc. Natl. Acad. Sci.* 97(6), 2486-2490 (2000).

Huang, B.; Guo, J.; Yi, B.; Yu, X.; Sun, L.; and Chen, W. Heterologous production of secondary metabolites as pharmaceuticals in *Saccharomyces cerevisiae*. *Biotechnol. Lett.* 30, 1121–1137 (2008).

Jardine, O.; Gough, J.; Chothia, C.; and Teichmann, S. A. Comparison of the small molecule metabolic enzymes of *Escherichia coli* and *Saccharomyces cerevisiae*. *Genome Res.* 12, 916-929 (2002).

Justino, M. C.; Almeida, C. C.; Teixeira, M.; and Saraiva, L. M. *Escherichia coli* di-iron YtfE protein is necessary for the repair of stress-damaged iron-sulfur clusters. *J. Biol. Chem.* 282(14), 10352–10359 (2007).

Kim, H.; Kim, H.; Kim, J.; Ryu, K.; Park, S.; Jahng, K.; Yang, M.; and Kim, D. Expression of heteropolymeric ferritin improves iron storage in *Saccharomyces cerevisiae*. *Appl. Environ. Microb.* 69, 1999-2005 (2003).

Lange, B. M.; and Croteau, R. Isopentenyl diphosphate biosynthesis via a mevalonate-independent pathway: Isopentenyl monophosphate kinase catalyzes the terminal enzymatic step. *Proc. Natl. Acad. Sci.* 96(24), 13714-13719 (1999).

Lange, H.; Kaut, A.; Kispal, G.; and Lill, R. A mitochondrial ferredoxin is essential for biogenesis of cellular iron-sulfur proteins. *Proc. Natl. Acad. Sci.* 97(3), 1050-1055 (2000).

Lill, R.; and Mühlenhoff, U. Maturation of iron-sulfur proteins in Eukaryotes: mechanisms, connected processes, and diseases. *Annu. Rev. Biochem.* 77, 669-700 (2008).

Loiseau, L.; Gerez, C.; Martijn, B.; Chouden, S. O.; Py, B.; Sanakis, Y.; de Mattos, J. T.; Fontecave, M.; and Barres, F. ErpA, an iron-sulfur (Fe-S) protein of the A-type essential for respiratory metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 104(34), 13626-13631 (2007).

Lüttgen, H.; Rohdich, F.; Herz, S.; Wungsintaweekul, J.; Hecht, S.; Schuhr, C. A.; Fellermeier, M.; Sagner, S.; Zenk, M. H.; Bacher, A.; and Eisenreich, W. Biosynthesis of terpenoids: YchB protein of *Escherichia coli* phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-d-erythritol. *Proc. Natl. Acad. Sci.* 97(3), 1062-1067 (2000).

Madsen, K. M.; Udatha, G. D. R. K.; Semba, S.; Otero, J. M.; Kötter, P.; Nielsen, J.; Ebizuka, Y. Kushiro, T.; and Panagiotou, G. Linking genotype and phenotype of *Saccharomyces cerevisiae* strains reveals metabolic engineering targets and leads to triterpene hyper-producers. *Plos One.* 6(3), e14763 (2011).

Mansy, S.S.; Wu, G.; Surerus, K.K.; and Cowan, J.A. Iron-sulfur cluster biosynthesis. *Thermatoga maritima* IscU is a structured iron-sulfur cluster assembly protein. *J. Biol. Chem.* 277, 21397-21404 (2002).

Maimone, T. J.; and Baran, P. S. Modern synthetic efforts towards biologically active terpenes. *Nat. Chem. Biol.* 3(7), 396-407 (2007).

Martin, J. J. V.; Pitera, D. J.; Withers, S. T.; Newman J. D.; and Keasling, J. D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21(7), 796-802 (2003).

Maury, J.; Asadollahi, M. A.; Møller, K.; Schalk, M.; Clark, A.; Formenti, L. R.; and Nielsen, J. Reconstruction of a bacterial isoprenoid biosynthetic pathway in *Saccharomyces cerevisiae*. *FEBS Lett.* 582, 4032-4038 (2008).

Mumberg, D.; Müller, R.; and Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene.* 156(1), 119-122 (1995).

Mühlenhoff, U.; and Lill, R. Biogenesis of iron-sulfur proteins in eukaryotes: a novel task of mitochondria that is inherited from bacteria. *Biochim. Biophys. Acta.* 1459, 370-382 (2000).

Nakamura, M.; Saeki, K.; and Takahashi, Y. Hyperproduction of recombinant ferredoxins in *Escherichia coli* by coexpression of the ORF1-ORF2-*iscS-iscU-iscA-hscB-hscA-fdx*-ORF3 gene cluster. *J. Biochem.* 126, 10-18 (1999).

Nielsen, K. F.; and Madsen, J. Ø. Determination of ergosterol on mouldy building materials using isotope dilution and gas chromatography–tandem mass spectrometry. *J. Chromatogr. A*, 898, 227–234 (2000).

Pitera, D. J.; Paddon, C. J.; Newman, J. D.; and Keasling, J. D. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.* 9, 193-207 (2007).

Puan, K.; Wang, H.; Dairi, T.; Kuzuyama, T.; and Morita, C. *fldA* is an essential gene required in the 2-C-methyl-D-erythritol 4 phosphate pathway for isoprenoid biosynthesis. *FEBS lett.* 579, 3802-3806 (2005).

Py, B.; and Barras, F. Building Fe-S proteins: bacterial strategies. *Nature reviews.* 8, 436-446 (2010).

Raguzzi, F.; Lesuisse, E.; and Crichton, R. R. Iron storage in *Saccharomyces cerevisiae*. *FEBS Lett.* 231(1), 253-258 (1988).

Rohdich, F.; Wungsintaweeikul, J.; Fellermeier, M.; Sagner, S.; Herz, S.; Kis, K.; Eisenreich, W.; Bacher, A.; and Zenk, M. H. Cytidine 5'-triphosphate-dependent biosynthesis of isoprenoids: YgbP protein of *Escherichia coli* catalyzes the formation of 4-diphosphocytidyl-2-C-methylerythritol. *Proc. Natl. Acad. Sci.* 96(21), 11758-11763 (1999).

Rohdich, F.; Hecht, S.; Gärtner, K.; Adam, P.; Krieger, C.; Amslinger, S.; Arigoni, D.; Bacher, A.; and Eisenreich, W. Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. *Proc. Natl. Acad. Sci.* 99(3), 1158-1163 (2002).

Seemann, M.; Bui, B. T. S.; Wolff, M.; Tritscg, D.; Campos, N.; Boronat, A.; Marquet, A.; and Rohmer, M. Isoprenoid biosynthesis through the methylerythritol phosphate pathway: the (E)-4-Hydroxy-3-methylbut-2-enyl diphosphate synthase (GcpE) is a [4Fe–4S] protein. *Angew. Chem.* 114(22), 4513–4515 (2002).

Seemann, M.; and Rohmer, M. Isoprenoid biosynthesis via the methylerythritol phosphate pathway: GcpE and LytB, two novel iron sulphur proteins. *C. R. Chimie.* 10, 748-755 (2007).

Sheff, M. A.; and Thorn, K. S. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast.* 21, 661 – 670 (2004).

Skala, J.; Capieaux, E.; Balzi, E.; Chen, W.; and Goffeau, A. Yeast sequencing reports. Complete sequence of the *Saccharomyces cerevisiae* *LEU1* gene encoding isopropylmalate isomerase. *Yeast.* 7(3), 281–285 (1991).

Takahashi, S.; Kuzuyama, T.; Watanabe, H.; and Seto, H. A 1-deoxy-d-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-d-erythritol 4-phosphate in an

alternative nonmevalonate pathway for terpenoid biosynthesis. *Proc. Natl. Acad. Sci.* 95(17), 9879-9884 (1998).

Tokumoto, U.; Nomura, S.; Minami, Y.; Mihara, H.; Kato, S.; Kurihara, T.; Esaki, N.; Kanazawa, H.; Matsubara, H.; and Takahashi, Y. Network of protein-protein interactions among iron-sulfur cluster assembly proteins in *Escherichia coli*. *J. Biochem.* 131, 713-719 (2002).

Truan, G.; Cullin, C.; Reisdorf, P.; Urban, P.; and Pompon, D. Enhanced *in vivo* monooxygenase activities of mammalian P450s in engineered yeast cells producing high levels of NADPH-P450 reductase and human cytochrome b₅. *Gene.* 125, 49-55 (1993).

Tokumoto, U.; Kitamura, S.; Fukuyama, K.; and Takahashi, Y. Interchangeability and distinct properties of bacterial Fe-S cluster assembly systems: functional replacement of the *isc* and *suf* Operons in *Escherichia coli* with the *nifSU*-like operon from *Helicobacter pylori*. *J. Biochem.* 136, 199-209 (2004).

Umebayashi, K.; and Nakano, A. Ergosterol is required for targeting of tryptophan permease to the yeast plasma membrane. *J. Cell Biol.* 161(6), 1117-1131 (2003).

Veen, M. and Lang, C. Production of lipid compounds in the yeast *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 63, 635-646 (2004).

Winston, F.; Dollard, C.; and Ricupero-Hovasse, S. L. Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast.* 11, 53-55 (1995).

Xiang, S.; Usunow, G.; Lange, G.; Busch, M.; and Tong, L. Crystal structure of 1-deoxy-d-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. *J. Biol. Chem.* 282, 2676-2682 (2007).

Yoon, T.; and Cowan, J. A. Iron-sulfur cluster biosynthesis. Characterization of frataxin as an iron donor for assembly of [2Fe;2S] clusters in ISU-type proteins. *J. Am. Chem. Soc.* 125, 6078-6084 (2003).

Zepeck, F.; Gräwert, T.; Kaiser, J.; Schramek, N.; Eisenreich, W.; Bacher, A.; and Rohdich, F. Biosynthesis of Isoprenoids. Purification and properties of IspG protein from *Escherichia coli*. *J. Org. Chem.* 70, 9168-9174 (2005).

Zhou, H.; Cheng, J.-S.; Wang, B.; Fink, G. R.; and Stephanopoulos, G. Xylose isomerase overexpression along with engineering of the pentose phosphate pathway and evolutionary engineering enable rapid xylose utilization and ethanol production by *Saccharomyces cerevisiae*. *Metab. Eng.* (2012), <http://dx.doi.org/10.1016/j.ymben.2012.07.011>

Zhou, K.; Zou, R.; Stephanopoulos, G.; and Too, H.-P. Metabolite profiling identified methylerythritol cyclodiphosphate efflux as a limiting step in microbial isoprenoid production. PLOS One (in press).

2.7 Supplementary material

2.7.1 Supplementary figure legends

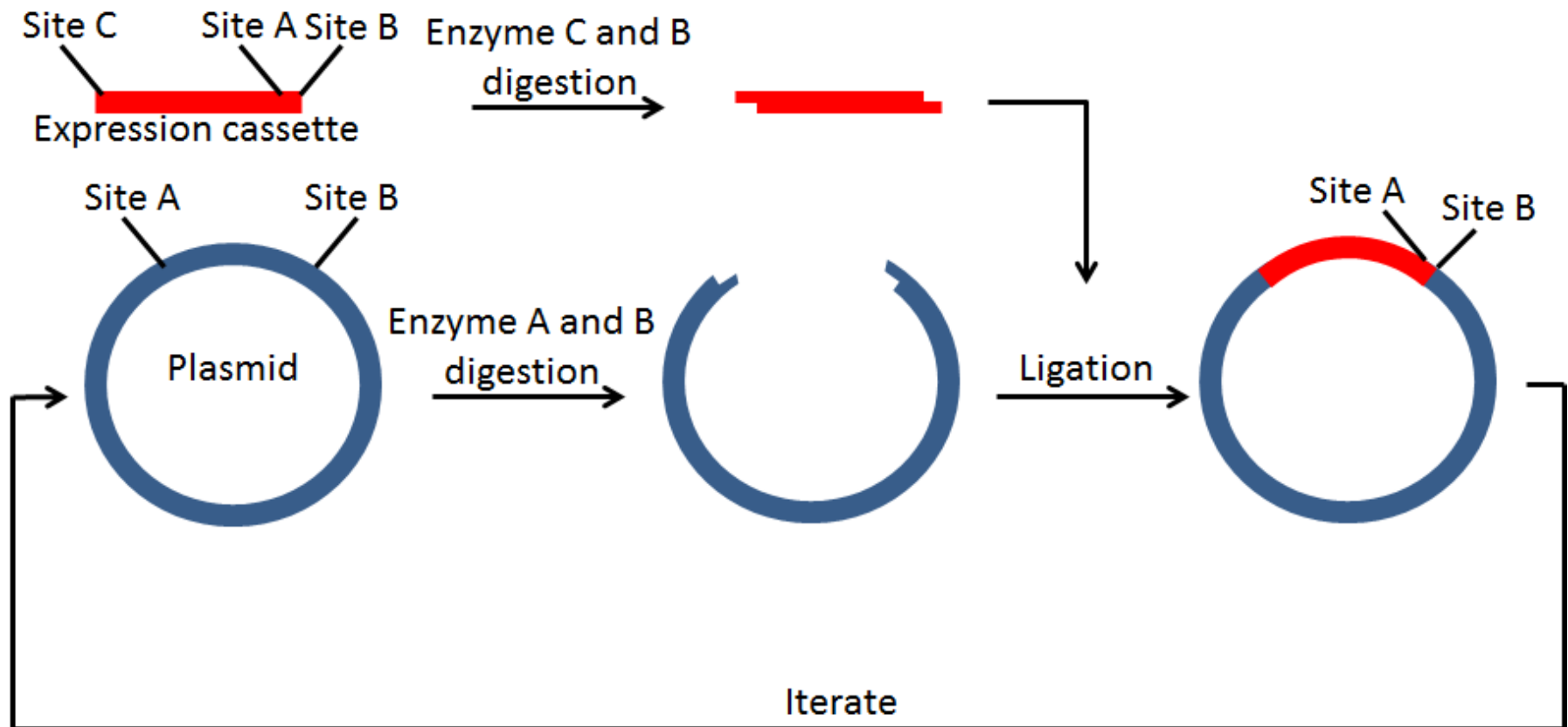
Supplementary Figure 2.1: Restriction endonucleases AvrII (A) and NheI (C) create compatible sticky ends, which upon ligation with each other cannot be redigested with either of the two enzymes. A cleavage site for restriction endonuclease NotI is indicated by B. DNA is amplified with primers containing the appropriate cleavage sites, treated with C and B, and ligated into a plasmid opened by A and B. This creates a plasmid with a new unique site A, maintaining site B and thus enabling insertion of another PCR product treated with C and B. This method can be iteratively used to assemble DNA as long as the restriction sites remain unique.

Supplementary Figure 2.2: Maps of plasmids applied in this study.

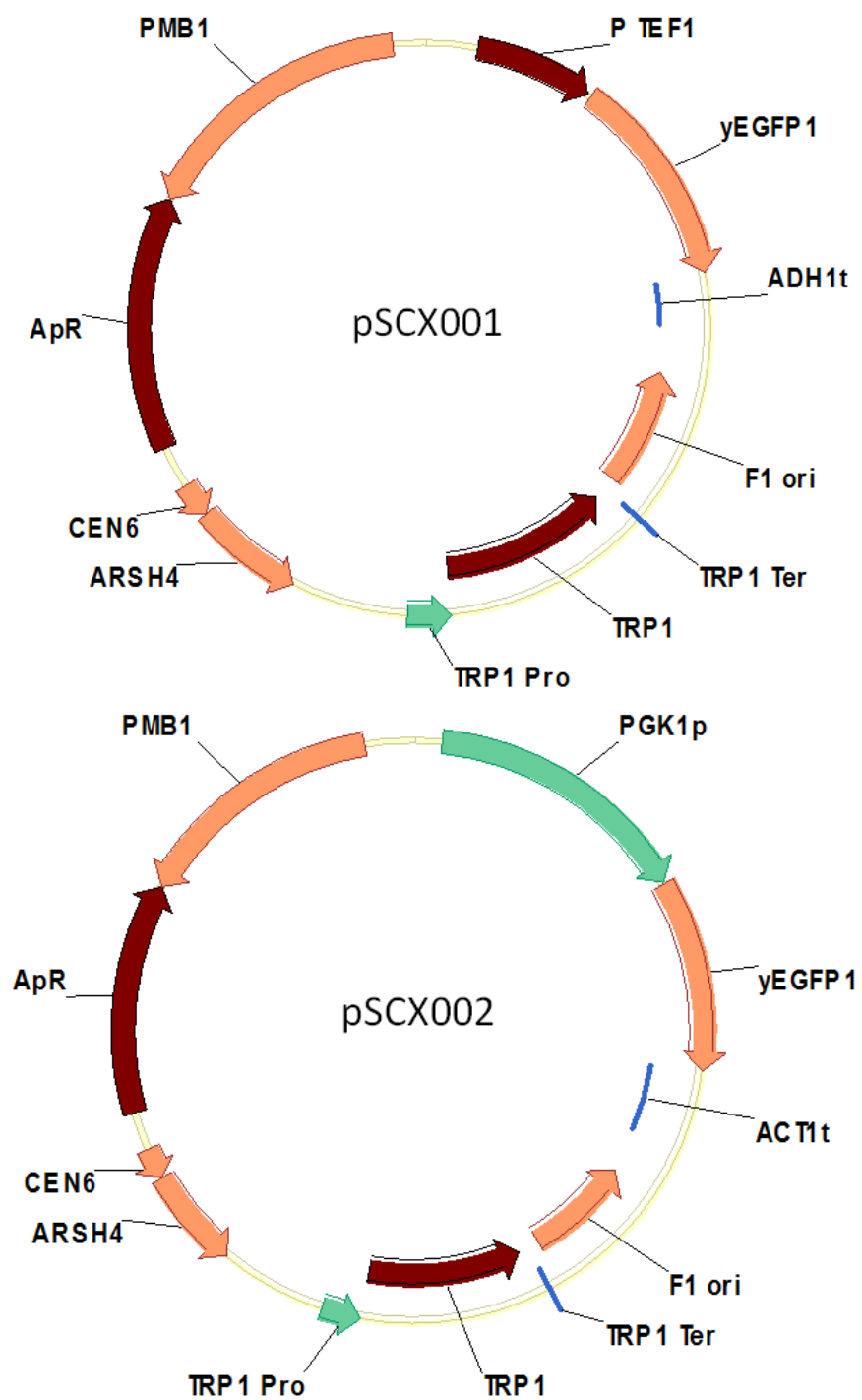
Supplementary Figure 2.3: GC-MS data. A: Total ion-count chromatogram of *S. cerevisiae* cell extract. B: MS-spectrum of ergosterol extracted from SCMEP-FeS-IS- Δ ERG13 grown on unlabeled glucose. C: MS-spectrum of ergosterol extracted from SCMEP-FeS-IS- Δ ERG13 grown on U- $^{13}\text{C}_6$ glucose.

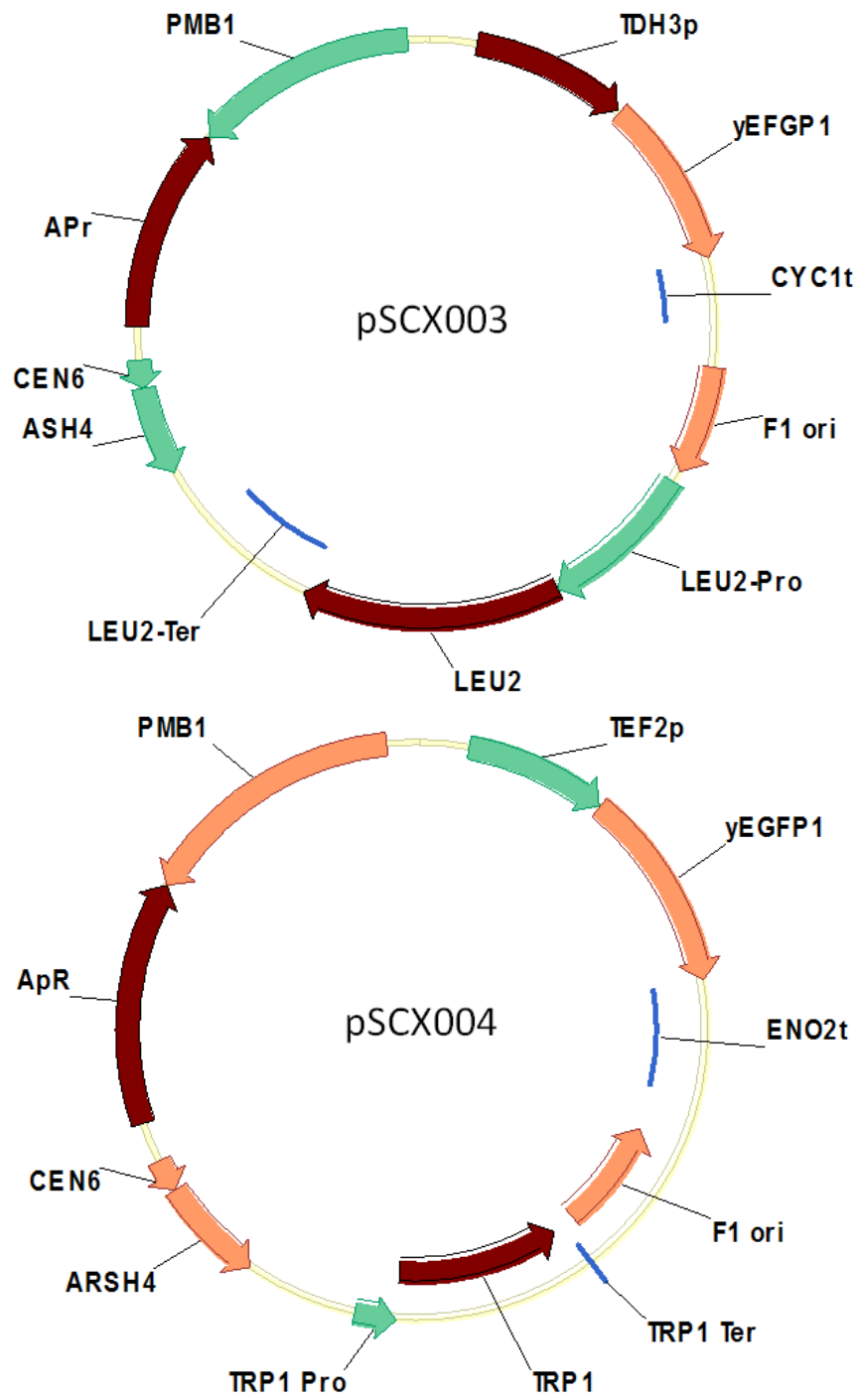
2.7.2 Supplementary figures

Supplementary Figure 2.1

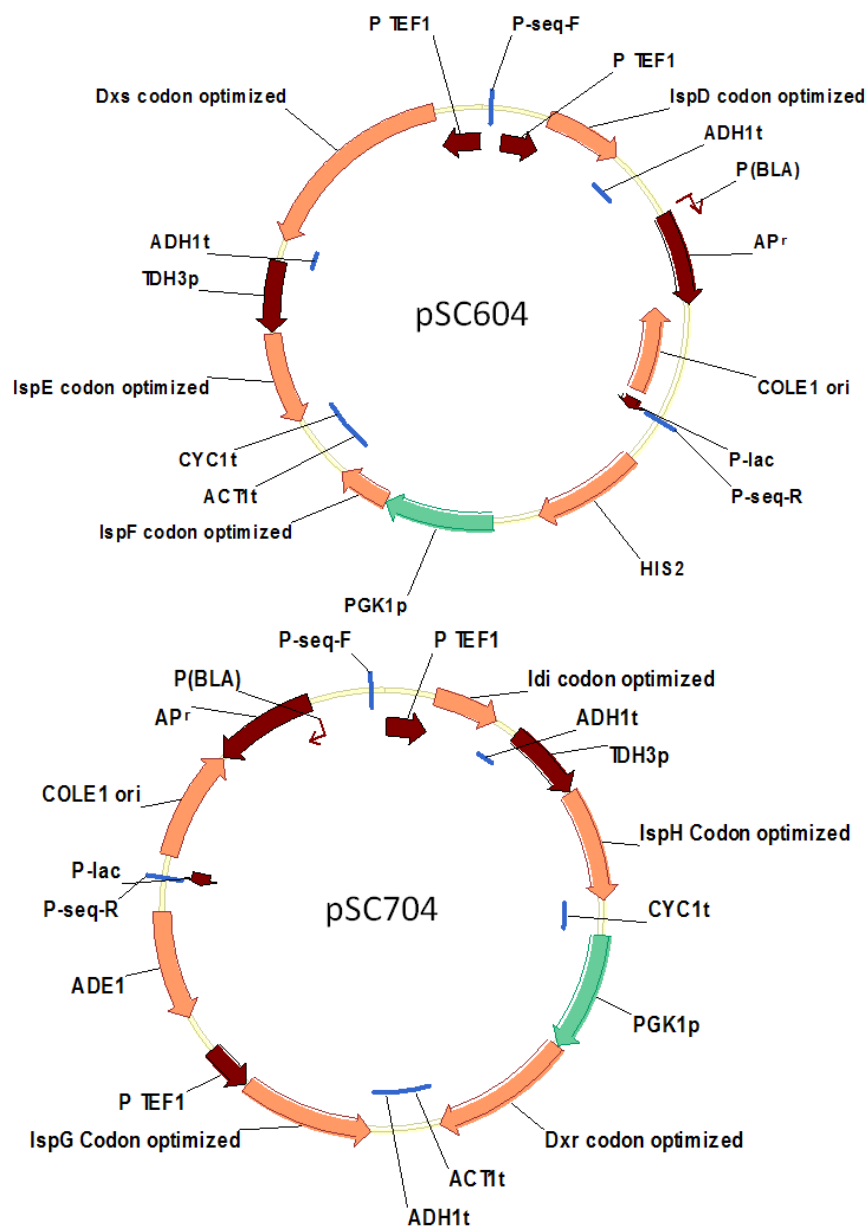


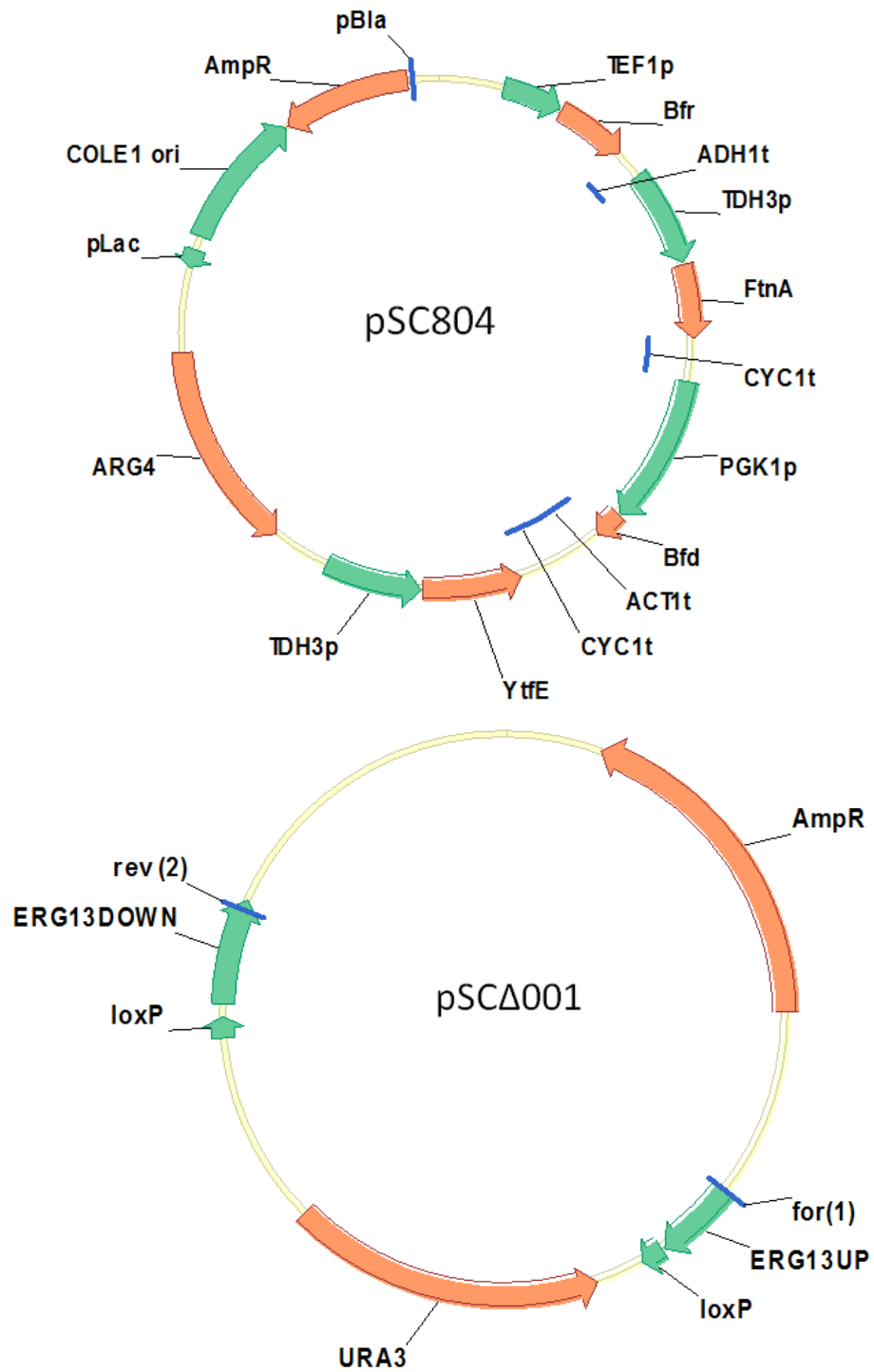
Supplementary Figure 2.2: Plasmid maps



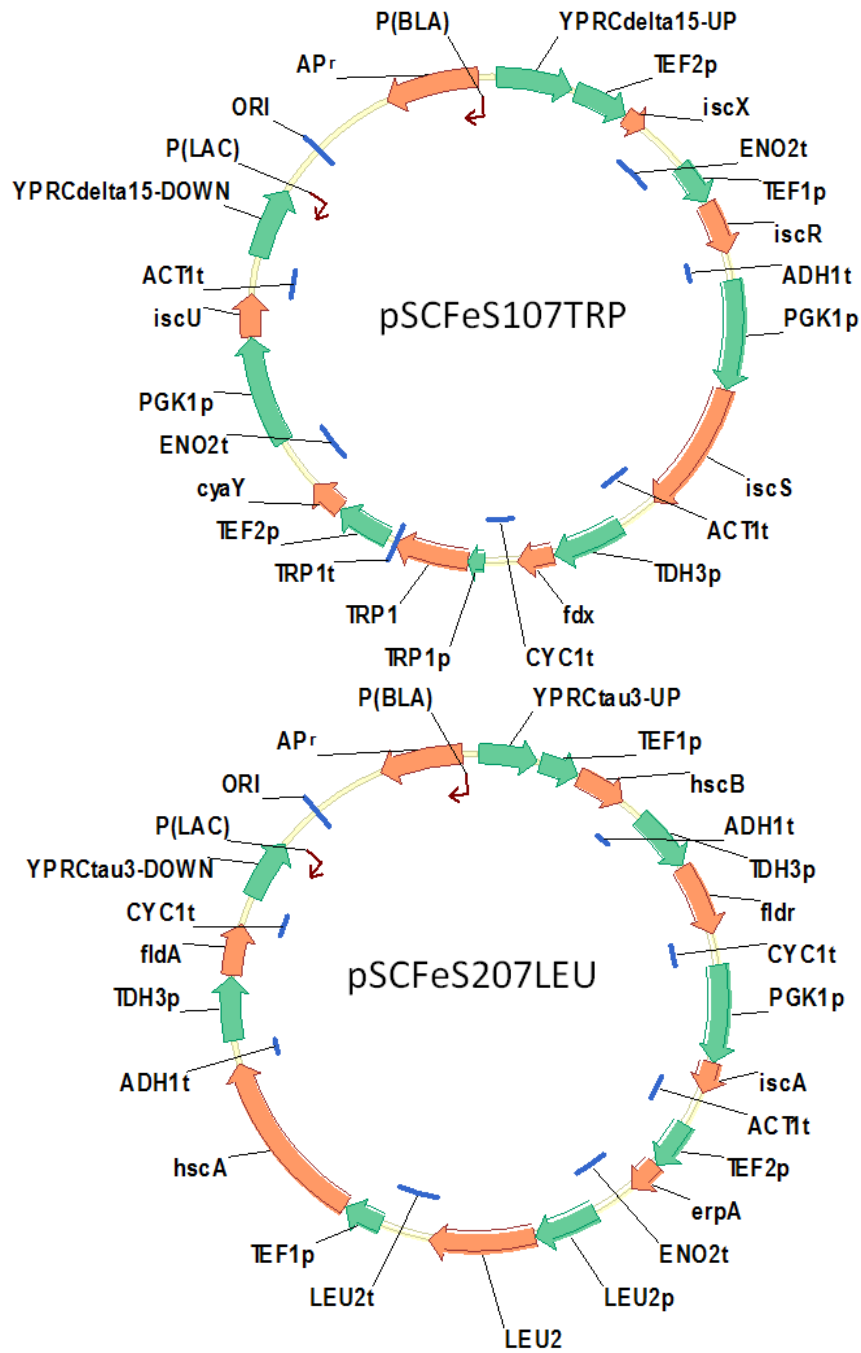


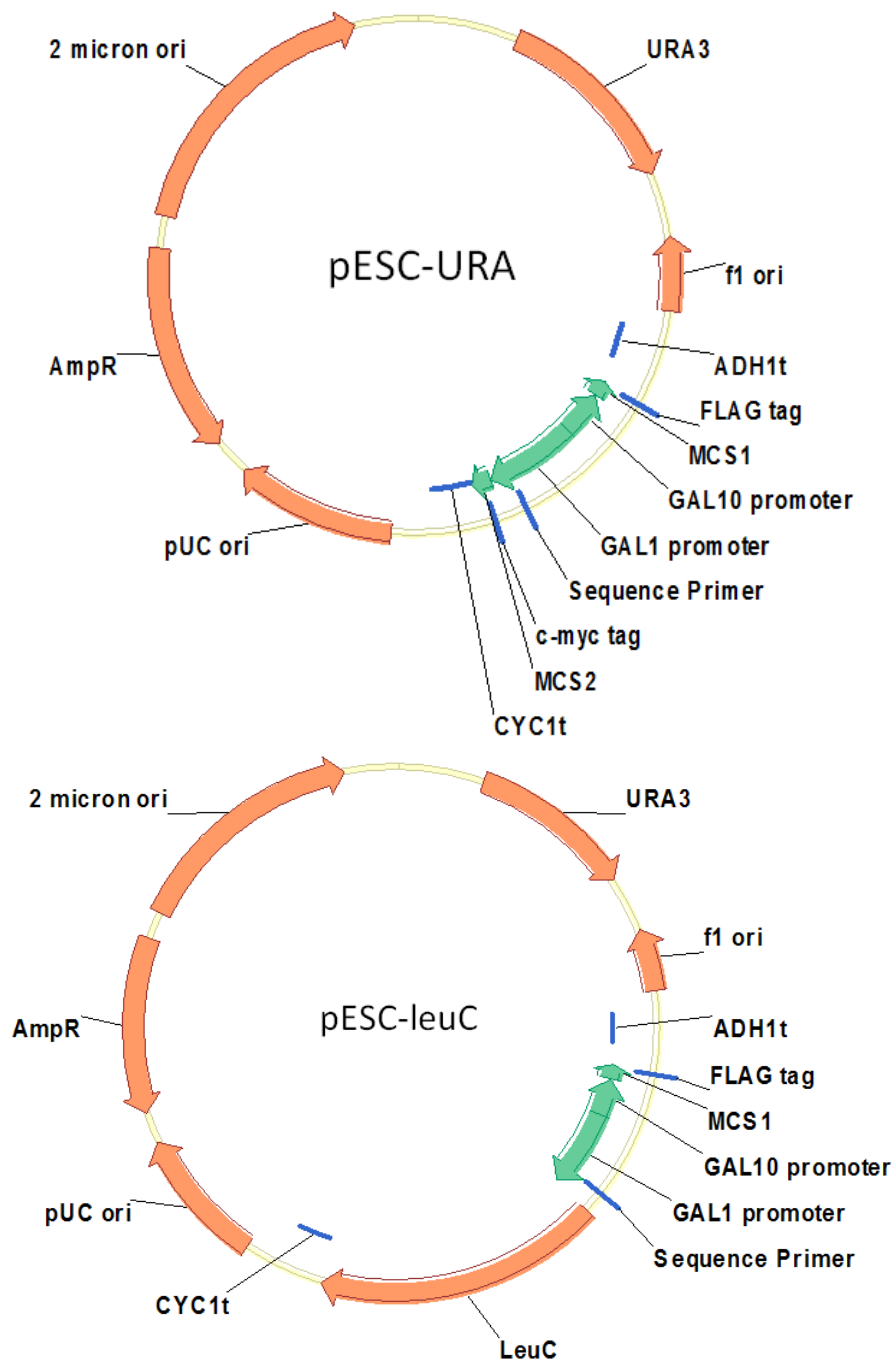
Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast



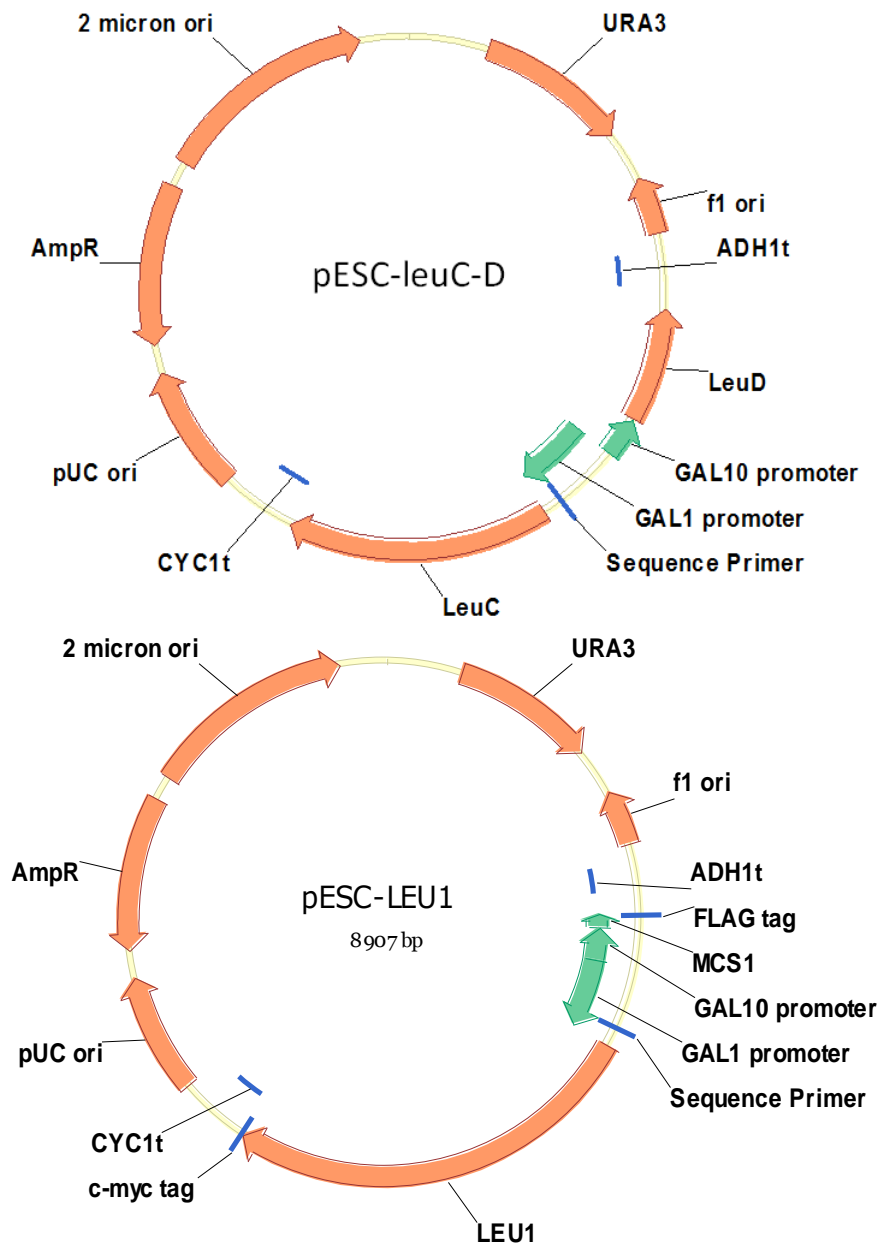


Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast

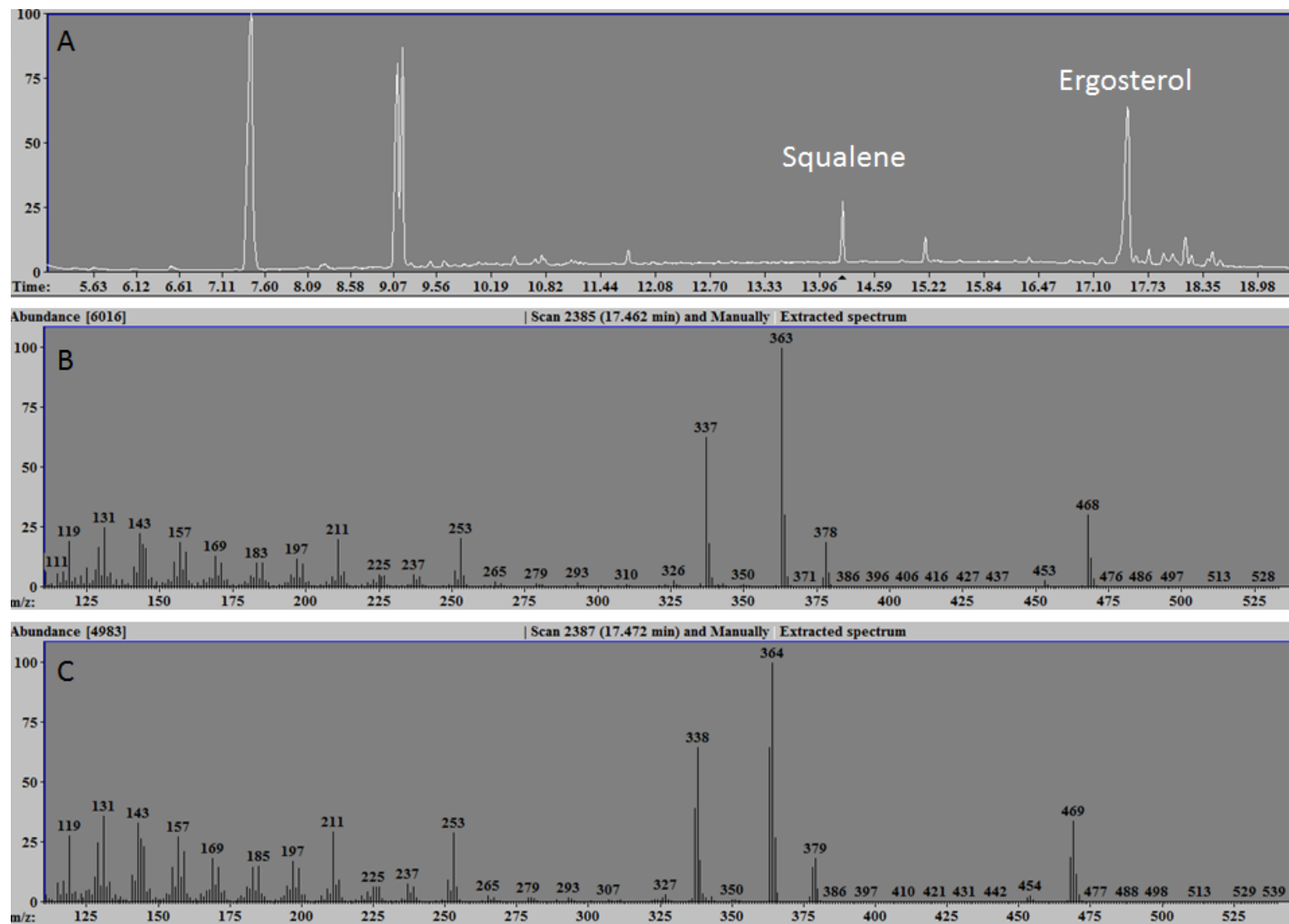




Expression of chromosomally integrated genes for the bacterial 2-C-methyl-D-erythritol 4-phosphate pathway in *Saccharomyces cerevisiae*: challenges of heterologous expression of bacterial iron-sulfur cluster genes in yeast



Supplementary Figure 2.3: GC-MS data.



Chapter 3 – Inverse metabolic engineering for increased production of isoprenoids in *Saccharomyces cerevisiae*: construction and screening of genomic DNA libraries

3.1 Abstract

Heterologous production of isoprenoids in microbes through fermentation is a promising alternative to isolation of the isoprenoid molecules from their native hosts since their scarce availability therein makes their purification both laborious and costly. Microbes are not naturally optimized for production of isoprenoids and therefore have to be optimized to improve their isoprenoid production capabilities. Instead of having to separately engineer a production system for each individual isoprenoid it would be beneficial to establish a microorganism as an isoprenoid production platform producing high amounts of the universal isoprenoid precursor isopentenyl pyrophosphate (IPP). In such a microbial production platform one would simply have to insert the genes encoding the downstream pathway converting IPP to the isoprenoid of interest followed by minor adjustments to obtain an optimized production system.

The methodology of metabolic engineering has been developed to enable the construction of microorganisms with desired phenotypes through direct genetic engineering. However, prior knowledge of the microbial system chosen for heterologous isoprenoid production is a prerequisite for the success of any direct metabolic engineering approaches. *Saccharomyces cerevisiae* holds great promise as a starting point for developing a microbial isoprenoid production platform due to all the available knowledge of its physiology, genetics and fermentation characteristics. Even though this great amount of knowledge is available for *S. cerevisiae*, very few genetic targets for increasing the flux towards IPP have been identified. To remedy this, we decided to employ inverse metabolic engineering in an effort to identify novel targets that could improve the isoprenoid precursor availability and thereby establish *S. cerevisiae* as a superior microbial isoprenoid production platform.

In this study, we sought potential overexpression targets from *S. cerevisiae* itself and pursued these through creation of a constitutive overexpression gDNA library from CEN.PK113-9D gDNA based on the pRS415TDH3 plasmid. We were able to create a gDNA library where 1/10 of the total library, when transformed into *E. coli*, resulted in 100,000 CFU of which 80% contained an insert. The library was screened for its ability to improve the heterologous production in *S. cerevisiae* of lycopene, an isoprenoid surrogate compound whose production correlates with increased flux through the isoprenoid biosynthetic pathway. As a red pigment lycopene allowed us to visually screen the library through the selection of colonies with increased red coloration. 200 colonies were picked for a second round of screening, and from these 200, 27 colonies were selected for testing in liquid culture. Plasmids from the 10 cultures producing the highest biomass yield of lycopene were isolated and transformed back into a clean background strain. Of the 10 plasmids only a single one could recover the lycopene production level found in the initial test, and when this plasmid was sequenced, it was found not to contain an open reading frame. The results highlight the feasibility of our approach but also that a greater number of colonies have to be analyzed in order to successfully identify plasmid-based lycopene overproducing mutants.

3.2 Introduction

With more than 55,000 identified compounds, isoprenoids constitute the largest group of secondary metabolites to date. This ubiquitous group of compounds holds great potential for identification of novel bioactive molecules. Several isoprenoids such as paclitaxel and artemisinin have already been demonstrated to be applicable as anticancer and antimalarial drugs, respectively. An inherent problem in production of isoprenoids is the limited availability of the molecules within their natural organism of production, which renders their production costly and labor intensive [Ajikumar et al., 2008; Muntendam et al., 2009].

As an alternative to isolating isoprenoids from their natural sources, heterologous microbial production holds great potential. By applying direct metabolic engineering, the gene(s) involved in the biosynthesis of a given molecule can be transferred to a well-characterized microbial host. The advantages of a microbial host, compared to natural isoprenoid-producing plants or higher eukaryotes, include higher productivity due to rapid cell growth and independence of seasonal variations caused by environmental stimuli e.g. weather conditions and nutrient availability. Furthermore microorganisms can be grown on relative cheap substrates in large fermenters where rigid process control allows for product quality consistency [Chemler and Koffas, 2008; Lee et al., 2009]. The advantage of applying well characterized model organism such as *Escherichia coli* and *S. cerevisiae* over microbes naturally producing isoprenoids of interest is the ease of which the model organisms can be genetically engineered to further improve production and the well-established fermentation protocols allowing for easy scale up and production.

S. cerevisiae has been applied in fermentative production for centuries, which has resulted in the vast accumulation of knowledge about its genetics, physiology, and genetic engineering. Its applicability is further supported by having a fully-sequenced genome and by the existence of molecular biology tools such as selectable markers, expression vectors, affinity tags and reporter genes. Most of this information is made readily accessible in databases such as Saccharomyces Genome Database (<http://www.yeastgenome.org>) and Comprehensive Yeast Genome Database (<http://mips.gsf.de/genre/proj/yeast/>) [Nevoigt, 2009]. Even though *E. coli* has only been applied in biotechnological production for a very short time relative to *S. cerevisiae*, the knowledge and tools available for this systems rivals that of *S. cerevisiae*. *E. coli* has a higher growth rate, which results in higher turnover rates, whereas *S. cerevisiae* has greater tolerance towards high sugar concentrations and low pH, minimizing the risk of contamination [Porro et al., 2011]. Of great importance for production of several isoprenoid molecules is the oxidative decoration catalyzed by cytochrome P450. For functional expression of cytochrome P450, it needs to be linked to a reductase, either transcriptionally or by co-localization. The disadvantage of transcriptional linking is that it has to be done individually in each case, whereas co-localization allows for a single reductase to link with multiple cytochrome P450s. In the latter case, a strain overexpressing a reductase targeted to a specific compartment in the cell, such as an organelle,

can be constructed thereby serving as a platform for heterologous cytochrome P450 expression. As a eukaryotic organism, *S. cerevisiae* possesses organelles, making it superior to *E. coli* for production of isoprenoids that require cytochrome P450s [Kirby and Keasling, 2009]. In the end the best choice of production platform depends on the specific product. However, we have selected *S. cerevisiae* in which to identify novel genetic targets for improving isoprenoid production since an inferior number of targets for this organism have been identified so far compared to the other major organism applied for isoprenoid production, namely *E. coli*. This limited knowledge hampers the application of direct metabolic engineering of *S. cerevisiae* for isoprenoid production, which in turn hinders the construction of a strain capable of producing isoprenoids at levels that are cost-competitive with organic synthesis. We are convinced that *S. cerevisiae* holds great potential and given that novel targets are identified it will be possible through metabolic engineering to establish it as an isoprenoid production platform. For a short review on the engineering efforts to increase the isoprenoid production within the two organisms see the tables in [Ajikumar et al., 2008].

One of the main contributions to strain improvement has come from creating diversity within a microbial population through mutagenesis followed by selection or screening. Prior to the advent of modern genetic engineering, the method known today as classical strain improvement was conducted through global random approaches such as chemical, UV-light, or X-ray induced mutagenesis [Adrio and Demain, 2006]. Classical strain improvement has several drawbacks such as accumulation of non-beneficial mutations and it is inherently difficult to identify the causative genotype of a desired phenotype. Today diversity within a microbial production can be created via introduction of DNA libraries e.g. gDNA and cDNA libraries. Using this approach, the genotype causing the phenotype for which the library has been screened can easily be traced back to the DNA originating from the library initially used to transform the population. This approach can be considered as a subsidiary of inverse metabolic engineering. Inverse metabolic engineering is defined as: “first, identifying, constructing, or calculating a desired phenotype; second, determining the genetic or the particular environmental factors conferring that phenotype; and third, endowing that phenotype on another strain or organism by directed genetic or environmental manipulation” [Bailey et al., 1996]. In general inverse metabolic engineering relies on two organisms with discrete phenotypes and can be seen as the search for the genetic element which can impose the desired phenotype of one organism on the other. In special cases within inverse metabolic engineering, the application of two distinct phenotypes can be neglected, e.g. DNA from the same organism can be inserted back into itself with altered copy number, expression level or other modification to identify overexpression targets and/or copy number effects [Ramer et al., 1992].

Imperative for the success of an inverse metabolic engineering approach relying on DNA libraries to create diversity is a high-throughput screen to identify the colonies with the desired phenotype. Several methods for screening libraries have been developed. The most effective screening method is known as selection. Selection relies on the desired phenotype exhibiting a

higher growth rate than the undesired. In the best case selection will be discrete, and only the mutants with a desired phenotype will survive (“cloning by complementation” can be seen as a sub-group of this selection strategy). In cases where the growth rates exist over a continuous range, the important parameter is the selection ratio, which is defined as

$$R_s = \frac{\mu_{mutant}}{\mu_{parent}}$$

If the selection ration is higher than 1.2, the selection can be carried out on solid media, whereas if it is lower than 1.2, selection in continuous cultures such as chemostat cultures is required [Zhang et al., 2006]. This type of inverse metabolic engineering strategy, among others, has been applied in order to identify a gene from a *Pichia stipites* gDNA library which could improve xylose assimilation in *S. cerevisiae*. The diversified population was enriched by serial sub-culturing, thereby exploiting the favorable selection ratio for the mutants with an improved xylose assimilation phenotype. From the enriched population 16 mutants were analyzed, and it was found that 10 of them harbored the *P. stipites* *XYL3* gene [Jin et al., 2005].

However, it is not always possible to identify a selection pressure for which the desired phenotype confers an advantage, especially not for secondary metabolites produced heterologously. Secondary metabolites are defined as metabolites non-essential for growth but assumed to confer an advantage to their natural production organism since they would otherwise have been lost during the course of evolution [Williams et al., 1989]. This advantage is seldom transferred to a heterologous production host since the external environment of the heterologous organism does not apply similar selection pressures. On the contrary, heterologous production of secondary metabolites confers a growth disadvantage since precursor metabolites otherwise used for biomass generation is re-directed to produce the secondary metabolite. So unless a growth strategy imposing a selection pressure on the heterologous host similar to the one giving the natural producer an advantage conferred by the specific secondary metabolite, a mean other than selection is required for obtaining mutants with the desired phenotype.

Screening can also be conducted through a quantifiable phenotype, and as with selection, two cases exist, namely screening for discrete or continuous phenotypes. Discrete phenotypes are preferred over the continuous since the mutants exhibiting the desired phenotype can be picked in an either/or manner. With the continuous phenotypes it can be difficult to assess subtle changes, and if the screening is done for a phenotype which is difficult to manually quantify, the screening process easily becomes subjective, e.g. individual color perception. Screening has traditionally been performed manually by visual inspections of mutants on solid medium, but has recently been improved by automation with the advent of micro titer technology and robotics. An example of the screening for a discrete phenotype was published by Weinstock et al. (1983), who constructed a special vector allowing them to screen a gDNA library for open reading frames (ORF) with the application of the *lacZ* gene. Only colonies harboring an ORF would result in

blue colony coloration. Similar approaches were applied by Gray et al. (1982) and Ruby et al. (1983). An example of a continuous case is the identification of mutants with improved antibiotics or growth-inhibitory compound production. These mutants can be screened by the area of the clearing zone which corresponds to the potency of the produced compound and its production level [Lin et al. 2001].

No selection strategy or method for isolating isoprenoid overproducing mutants of *S. cerevisiae* has, to our knowledge, been identified so far. However, the naturally occurring pigment lycopene, belonging to the isoprenoid subgroup called carotenoids, has been extensively applied in both direct and inverse metabolic engineering schemes. Lycopene, as all other isoprenoids, is biosynthesized from IPP, which through a polymerization reaction with 2 molecules of dimethylallyl pyrophosphate (DMAPP) is converted to farnesyl pyrophosphate (FPP). This reaction in *S. cerevisiae* is catalyzed by the FPP synthase encoded by the *ERG20* gene. FPP is reacted with another molecule of DMAPP to form geranylgeranyl pyrophosphate (GGPP) catalyzed by GGPP synthase, encoded by the *BTS1* gene in *S. cerevisiae* and by the *crtE* gene in bacteria. The last two steps of the pathway are not endogenously present in *S. cerevisiae* but can be introduced into yeast by genetic engineering of the bacterial *crtB* and *crtI* genes, encoding phytoene synthase and phytoene dehydrogenase, respectively. The biosynthetic scheme for lycopene can be seen in Figure 3.1.

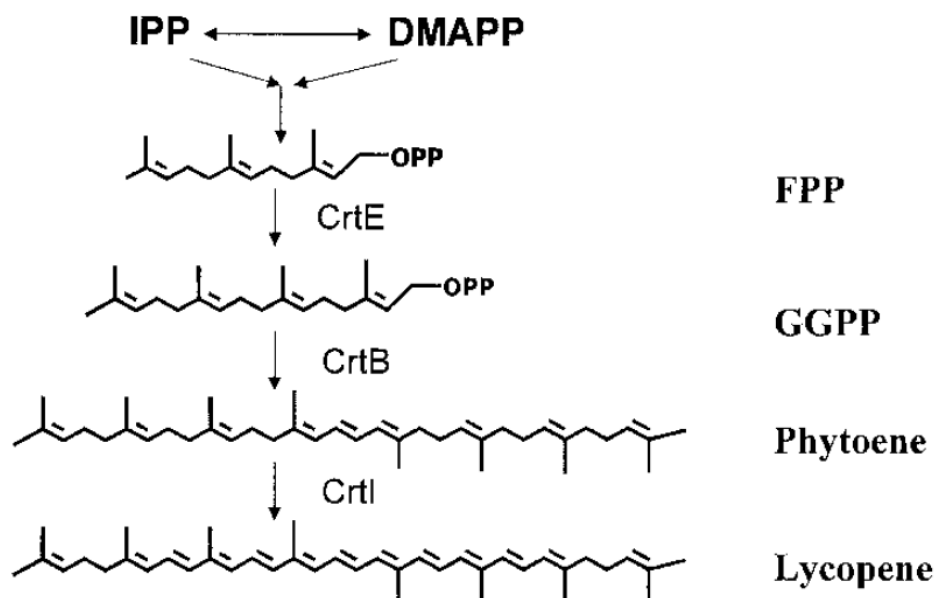


Figure 3.1: Lycopene biosynthetic pathway. Modified from Sedkova et al. (2005). IPP and two molecules of DMAPP are converted into FPP, catalyzed by the FPP synthase, and subsequently FPP is condensed with another molecule of DMAPP to form GGPP, catalyzed by GGPP synthase. Two molecules of GGPP are condensed head to head to form phytoene, catalyzed by the phytoene synthase. Phytoene is desaturated to form lycopene by the phytoene dehydrogenase. The enzymes catalyzing the four steps in our *S. cerevisiae* strain engineered to heterologously produce lycopene are encoded by the *ERG20*, *crtE*, *crtB* and *crtI* genes, respectively.

Lycopene has been applied as a surrogate or reporter metabolite in studies spanning from random mutagenesis of single protein targets [Leonard et al., 2010], over the effect of chromosomal mutations on plasmid copy number [Tao et al. 2005], stoichiometric modeling [Alper et al., 2005(a)], combinatorial effects of deletions and overexpression targets [Jin and Stephanopoulos, 2007], to the screening of gDNA- [Kang et al., 2005], cDNA- [Gallagher et al., 2003] and transposon mutagenesis- [Alper et al., 2005(b)], to global transcription factor engineering-libraries [Alper et al., 2007].

As mentioned above, lycopene has been especially successful as a surrogate metabolite applied in identifying novel genetic targets in *E. coli* for improving isoprenoid production. The method relied on cloning partially digested *E. coli* gDNA into the high copy number vector pBluescript to create a shot-gun library. This library was then transformed into *E. coli*, and the best strains were picked based on their colony coloration [Kang et al., 2005]. pBluescript is a very high copy

number vector in *E. coli*, present in 500-700 copies per chromosome [Mayer, 1995]. The improved lycopene production was therefore probably a result of the increased copy number of the identified genes. The application of a high copy vector in such an endeavor has several drawbacks: the effect of increased copy number will only have a marginal effect on tightly controlled genes, genes encoding products which are toxic at elevated levels cannot be identified, and the resultant phenotype can be the result of multiple genotypes if more than one plasmid have transformed (and can be co-maintained within) a single cell [Ramer et al., 1992]. To circumvent this, a low copy vector can be used in combination with an inducible promoter. This was done by Ramer et al. (1992), who applied an YCp plasmid and the *GALI* promoter to construct an inducible expression library from *S. cerevisiae* gDNA. For improvement of a secondary metabolite production phenotype, toxic genes are not of interest since they are non-applicable in most metabolic engineering strategies for industrial scale production. So to eliminate these, a strong constitutive promoter could be applied instead of the inducible. This approach could be a great way of identifying within *S. cerevisiae*'s genome novel targets that could be engineered to increase its ability to produce isoprenoids. If novel targets can be identified by this approach, they can subsequently serve as starting points in a metabolic engineering effort to establish *S. cerevisiae* as an isoprenoid production platform through a combinatorial gene expression approach. The approach could also be applied to other organisms to identify gDNA fragments which upon expression in *S. cerevisiae* will improve isoprenoid production.

We here demonstrate the construction of a constitutive expression library based on *S. cerevisiae* gDNA cloned into an YCp plasmid and the subsequent application of the library to create diversity within a *S. cerevisiae* population pre-engineered to produce lycopene. The resultant transformants are screened for isoprenoid over-producing mutants by visual selection of mutants exhibiting increased red colony coloration due to increased production of lycopene. The general strategy for our approach can be seen in Figure 3.2.

**Inverse metabolic engineering for increased production of isoprenoids in *Saccharomyces cerevisiae*:
construction and screening of genomic DNA libraries**

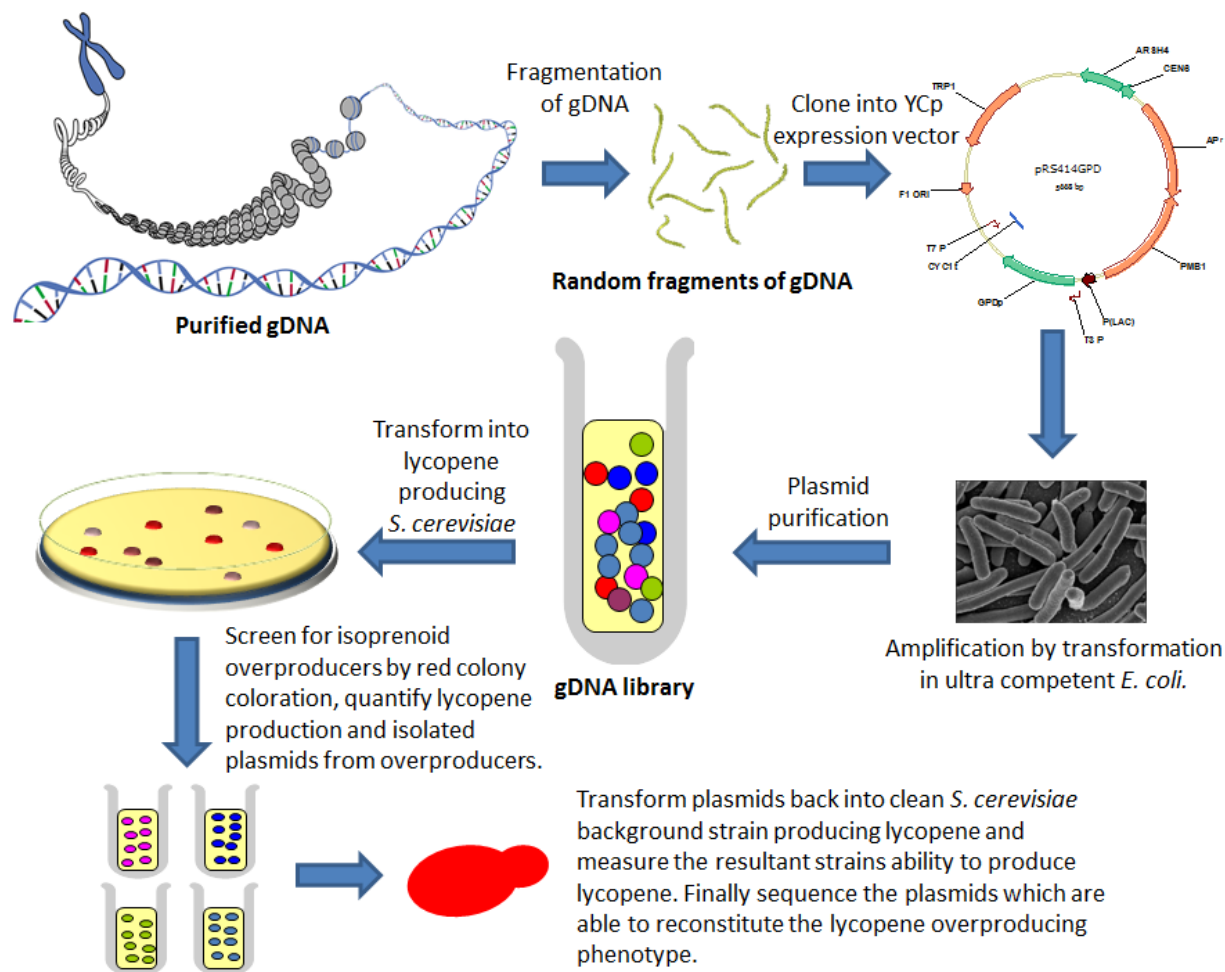


Figure 3.2: Strategy for construction and screening of gDNA libraries for lycopene overproducing mutants.

3.3 Materials and methods

3.3.1 Enzymes, chemicals and media

Unless otherwise stated, all enzymes applied in the construction of plasmids were purchased from New England Biolabs Inc., Ipswich, MA, and all chemicals from Sigma-Aldrich®, St. Louis, MO. For auxotrophic selection *S. cerevisiae* were grown on SC medium with the appropriate nutrient(s) omitted. This medium was composed of 6.7 g/l Yeast Nitrogen Base (YNB) without amino acids, Sunrise Science Products, Inc., San Diego; CSM – amino acid dropout mixture (amount according to manufacturer's recommendation), Sunrise Science Products, Inc., and 20 g/l dextrose. Growth of *S. cerevisiae* where no selection was required was conducted on YPD medium composed of: 10 g/l Bacto™ yeast extract, BD Diagnostic Systems, Sparks, MD; 20 g/l Bacto™ peptone, BD Diagnostic Systems; and 20 g/l dextrose. *E. coli* cells were grown in BD Difco™ LB broth from BD Diagnostic Systems. For selection and maintenance of plasmids, ampicillin was added to a final concentration of 100 mg/l. Solid media was made by addition of 20 g/l BD Bacto™ Agar, BD Diagnostic Systems.

3.3.2 Plasmid and strain construction

Construction of the plasmids described below was done by treating the recipient plasmids with endonuclease(s), purification by gel-electrophoresis and recovery by gel extraction using the PureLink™ Quick Gel Extraction Kit, Life Technologies, Grand Island, NY, followed by ligation to the insert using T4 DNA ligase. The inserts were either prepared by removing it from the plasmid harboring it by endonuclease digestion followed by gel-electrophoresis and purification in the same way as the plasmid recipient or amplified by PCR, purified using the QIAquick PCR Purification Kit, Qiagen, Valencia, CA, treated with endonucleases and purified again using the QIAquick PCR Purification Kit, Qiagen. PCR amplifications were performed using the Phusion® High-Fidelity DNA Polymerase. All constructs were verified by sequencing. The DNA ladder used as size reference was the 1 kb DNA ladder from New England Biolabs Inc. (bands corresponds to the following DNA lengths: 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 8.0, 10.0 kb). All the plasmids and strains acquired for this study can be seen in Table 3.1. For expression of the genes of interest, three expression cassettes were created and tested for their ability to mediate transcription by verifying that they could express the yeast enhanced green fluorescent protein (*yeGFP*). This was done on the basis of the pRS414TEF and pRS415TDH3 plasmids in which the *yeGFP* amplified by PCR from pKT127 was inserted by *SpeI/XhoI* digestion followed by ligation. The pRS414TEF plasmid with the inserted *yeGFP* was modified by exchanging the *CYC1t* with the *ADH1t* and *ACT1t* by *XhoI/KpnI* digestion. The *TEF1p* was substituted with the *PGK1p* in the plasmid with the *ACT1t* by *SacI/SpeI* digestion. The replacement promoters and terminators were obtained by PCR from CEN.PK113-7D genomic DNA. This resulted in the construction of 3 expression cassettes harboring *yeGFP*, namely

TEF1p-yeGFP-ADH1t, *TDH3p (GPDp)-yeGFP-CYC1t* and *PGK1p-yeGFP-ACT1t* in the pSC011, pSC012, and pSC013 plasmids, respectively. The carotenogenic genes *CrtE*, *CrtB* and *CrtI* were purchased codon optimized from Entelechon GmbH, Regensburg, Germany, based on the following protein sequences from *Pantoea agglomerans*, *CrtE* (GGPP synthase) - GenBank: AAA24819.1, *CrtB* (phytoene synthase) - GenBank: AAA24821.1 and *CrtI* (phytoene dehydrogenase) - GenBank: AAA24820.1. To construct individual expression cassettes the *CrtE*, *CrtB*, and *CrtI* were amplified by PCR and inserted in the *SpeI/XhoI* treated pSC011, pSC013, and pSC012 respectively. The three expression cassette were then sequentially inserted into the pRS306 plasmid which was previously modified to harbor a new multiple cloning site (MCS) and to have its *AatII* and *ApaI* restriction sites removed by silent point mutations, yielding the pSC203 plasmid. The new MCS was comprised of two complementary primers which were annealed followed by digestion with *SacI* and *KpnI* and inserted in pRS306. The new MCS contained 9 restriction enzyme recognition sites, whereof the *ApaI*, *AatII*, *AvrII* and *AflIII* were used to insert the expression cassettes. The point mutations were performed using the QuickChange® Multi Site-directed Mutagenesis kit, Stratagene, La Jolla, CA.

All transformations of *S. cerevisiae* were performed using the lithium acetate and polyethylene glycol method according to [Gietz and Woods, 2002]. Auxotrophic selection of the transformants was done on SC medium with the appropriate nutrient(s) omitted. Verification of the constructed strains was done by extraction of gDNA using the Wizard® Genomic DNA Purification Kit, Promega, Madison, WI, followed by PCR amplification of all the integrated ORFs. CEN.PK113-9D was transformed with the pSC203 plasmid to construct strain SimC002. Prior to the transformation, the plasmid was linearized with the *StuI* endonuclease. SimC002 was transformed with the pRS414TDH3 to construct the SimC008. The constructed strains can be seen in Table 3.1.

Table 3.1 - Strains and plasmids. The table shows the acquired and constructed, strains and plasmids applied in this study.

Strain	Genotype	Source
<i>Saccharomyces cerevisiae</i>		
CEN.PK113-7D	<i>MATa MAL2-8^c SUC2</i>	Dr. P. Kötter, Frankfurt, Germany.
CEN.PK113-9D	<i>MATa MAL2-8^c SUC2 trp1 ura3</i>	Dr. P. Kötter, Frankfurt, Germany.
SimC002	<i>MATa MAL2-8^c SUC2 trp1 ura3::pSC203(TEF1p-CrtE-ADH1t PGK1p-CrtB-ACT1t TDH3p-CrtI- CYC1t URA3)</i>	This study.
SimC008	<i>MATa MAL2-8^c SUC2 ura3::pSC203 (TEF1p-CrtE-ADH1t PGK1p-CrtB- ACT1t TDH3p-CrtI-CYC1t URA3) – pRS415TDH3 (TRP1)</i>	This study.
<i>Escherichia coli</i>		
MAX Efficiency® DH5α™	F- φ80lacZΔM15 Δ(lacZYA-argF) U169 <i>recA1 endA1 hsdR17</i> (rk ⁻ , mk ⁺) <i>phoA supE44 λ⁻ thi⁻1 gyrA96 relA1</i>	Life Technologies, Grand Island, NY.
ElectroMAX™ DH10B™ T1 Phage-Resistant Competent Cells	F ⁻ <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) φ80lacZ ΔM15 ΔlacX74 <i>recA1 endA1</i> <i>araD139</i> Δ(<i>ara, leu</i>)7697 <i>galU galK λ⁻</i> <i>rpsL nupG tonA</i>	Life Technologies, Grand Island, NY.
Plasmid name		Origin/reference
pSC203		Center for Microbial Biotechnology plasmid collection, DTU†.
pSC011		Center for Microbial Biotechnology plasmid collection,

	DTU†.
pSC012	Center for Microbial Biotechnology plasmid collection, DTU†.
pSC013	Center for Microbial Biotechnology plasmid collection, DTU†.
pRS306	[Sikorski and Hieter, 1989]
pKT127	[Sheff and Thorn, 2004]
pRS414TEF	[Mumberg et al., 1995]
pRS415TDH3	[Mumberg et al., 1995]
pRS414TDH3	[Mumberg et al., 1995]

† Constructed by Simon Carlsen as part of his Ph.D. thesis work.

3.3.3 Library construction

The gDNA library was created by inserting gDNA fragments in the *SalI* restriction site which, in the YCp plasmid pRS414TDH3, is located between the *TDH3* promoter and the *CYC1* terminator. The gDNA was isolated from CEN.PK113-9D using the Wizard® Genomic DNA Purification Kit, Promega and fragmented by partial digestion with *Sau3AI* endonuclease. Prior to the digestion the *Sau3AI* was diluted 10 fold in diluent buffer A, New England Biolabs Inc. A dilution series of the endonuclease concentration was set up with each reaction having a final volume of 100 µl. Each digestion reaction was stopped by addition of a 30 µl of a 1:2 mix of 0.5 M EDTA, pH 8, and gel loading buffer, New England Biolabs Inc. Following fragmentation, the gDNA was selected for the desired size distribution by gel-electrophoresis on a 0.75 % agarose gel. Fragments from 1.5-10 kb were purified using the PureLink™ Quick Gel Extraction Kit. The purified fragmented gDNA was then treated with Klenow Fragment (3'→5' exo-) with the addition of dATP and dGTP (nucleotides were acquired from Life Technologies), converting the 4 bp overhang to a 2 bp overhang, followed by purification with the QIAquick PCR Purification Kit, Qiagen. The pRS414TDH3 plasmid was treated with *SalI*-HF endonuclease for 8 h, purified by gel-electrophoresis, and recovered using the PureLink™ Quick Gel Extraction Kit, Life Technologies. The linearized plasmid was treated with Klenow (-exo) fragment with addition of dCTP and dTTP (nucleotides were acquired from Life Technologies) to avoid self-ligation. Subsequently, the plasmid was purified with the QIAquick PCR Purification Kit, Qiagen and

ligated in an estimated 1:1 molar ratio (total DNA concentration = 30 ng/μl, total volume 50 μl) to the fragmented gDNA with the T4 DNA ligase by incubation at 16 °C overnight. The ligation mix was precipitated with the addition of 5 μl 3 M sodium acetate, pH 5.2 and 50 μl isopropanol, washed twice with 50 μl ice cold 80 % ethanol and re-dissolved in 5 μl nuclease free water, Qiagen. A half microliter of the library was transformed into ElectroMAX™ DH10B™ T1 Phage-Resistant Competent Cells and plated on large (150 mm in diameter) LB + ampicillin plates and incubated overnight at 37 °C. Sixteen colonies were picked onto new LB + ampicillin plates for analysis of the library quality. The rest of the colonies were scraped of the plates using 1 ml ice cold LB medium per plate and a Drigalski spatula. Plasmids from the collected cells were purified using the Plasmid Midi Kit, Qiagen and hereafter constituted the working subset of the library while the remainder of the concentrated ligation mix was stored at -20 °C as a master library in case more working library has to be generated. The quality of the library was tested by colony PCR to verify how many of the 16 colonies harbored a pRS414TDH3 plasmid with an inserted gDNA fragment. The protocol was inspired by Rose and Broach (1991).

3.3.4 Library screening

SimC002 was transformed with 5 μg of the working subset of the library using the lithium acetate and polyethylene glycol method according to Gietz and Woods (2002) and plated on 50 large (150 mm diameter) SC-TRP-URA plates. The plates were incubated at 30 °C for three days, and based on the intensity of the red colony coloration, 200 mutants were picked onto new SC-TRP-URA plates and incubated for 2 days at 30 °C. Out of the 200 colonies, 27 colonies were picked for further testing. The strains were tested in triplicate, each in 5 ml SC-TRP-URA, for their lycopene production levels. The procedure for quantification of lycopene production is described in the following section.

3.3.5 Lycopene extraction, verification and quantification

The CEN.PK113-7D (negative control) and SimC002 strains were tested for the ability to produce lycopene by growing them on 5 ml SC-TRP-URA in 14 ml Falcon 2059 tubes, BD Biosciences. The cultures were inoculated at OD₆₀₀ = 0.1 from pre-cultures and incubated at 30 °C with 250 rpm orbital shaking for 2 days. The OD₆₀₀ were measured spectrophotometrically, followed by harvesting of the cells from 1 ml of culture. The cultures were harvested by centrifugation at 3000 g for 10 min, washed with 1 ml phosphate-buffered saline (Cellgro®, Mediatech, Inc., Manassas, VA) and transferred to a 2 ml FastPrep® tube (MP Biomedicals LLC., Solon, OH). The cells were spun down at 18,000 g for 1 min, and the phosphate-buffered saline removed. Two hundred μl acid washed glass beads (425-600 μm, Sigma-Aldrich®) and 1 ml of acetone containing 100 mg/l 2,6-di-*tert*-butyl-4-methylphenol (BHT) was added to each FastPrep® tube. The cells were disrupted using a Savant FastPrep® FP120 Cell Disrupter (MP

Biomedicals LLC.), and the extracts were filtered using a 13 mm syringe filter (w/ 0.2 μ m PTFE filter, VWR, Radnor, PA). Finally the lycopene concentration was determined spectrophotometrically by measuring the absorbance of the extract at 474 nm. The absorbance values were converted to concentration using a standard curve made in triplicate from genuine lycopene. To verify that the red pigment was lycopene, the absorption spectra from the extract and lycopene standard in acetone with 100 mg/l BHT were recorded from 350-900 nm and compared. The same approach was applied to quantify the lycopene production levels of the isolated mutants.

3.4 Results and discussion

3.4.1 Strain construction and verification

The pSC203 plasmid harboring the carotenogenic genes encoding the lycopene biosynthesis pathway was used to transform the *S. cerevisiae* CEN.PK113-9D strain. The resultant transformant was picked and tested by PCR to verify that the genes had been inserted in the genome. The results confirmed that the genes were inserted (result not shown), and the strain was named SimC002. To construct the control strain, SimC002 was transformed with the pRS414TDH3 plasmid, which likewise was confirmed by PCR (results not shown). The resultant strain was named SimC008. SimC002 and SimC008 were tested for their ability to produce lycopene by extracting biomass into acetone and spectrophotometrically verifying that the absorption spectra recorded from 350-900 nm match the reference spectrum found in the Carotenoid Handbook [Britton et al., 2004] (result not shown).

3.4.2 Library construction

In order to construct a *S. cerevisiae* gDNA library, gDNA was purified from CEN.PK113-9D. The gDNA size and quality was tested by gel-electrophoresis. The result can be seen in Figure 3.3, lane 1, which shows that the obtained gDNA has a high concentration and integrity. To fragment the gDNA, the endonuclease Sau3AI was applied, and a dilution series of the Sau3AI was set up to determine the concentration of the enzyme which would give the best size range of partially digested gDNA. The results can be seen in Figure 3.3, lanes 4-14. The desired size range of partially digested gDNA was selected as that obtained in between the 4th and 6th dilutions, corresponding to lanes 8-10 in Figure 3.3. It is important to notice that the amount of endonuclease required to obtain the desired size distribution of fragmented gDNA depends both on the initial gDNA concentration and size, which means that the required endonuclease amount has to be determined experimentally for each batch of purified gDNA.

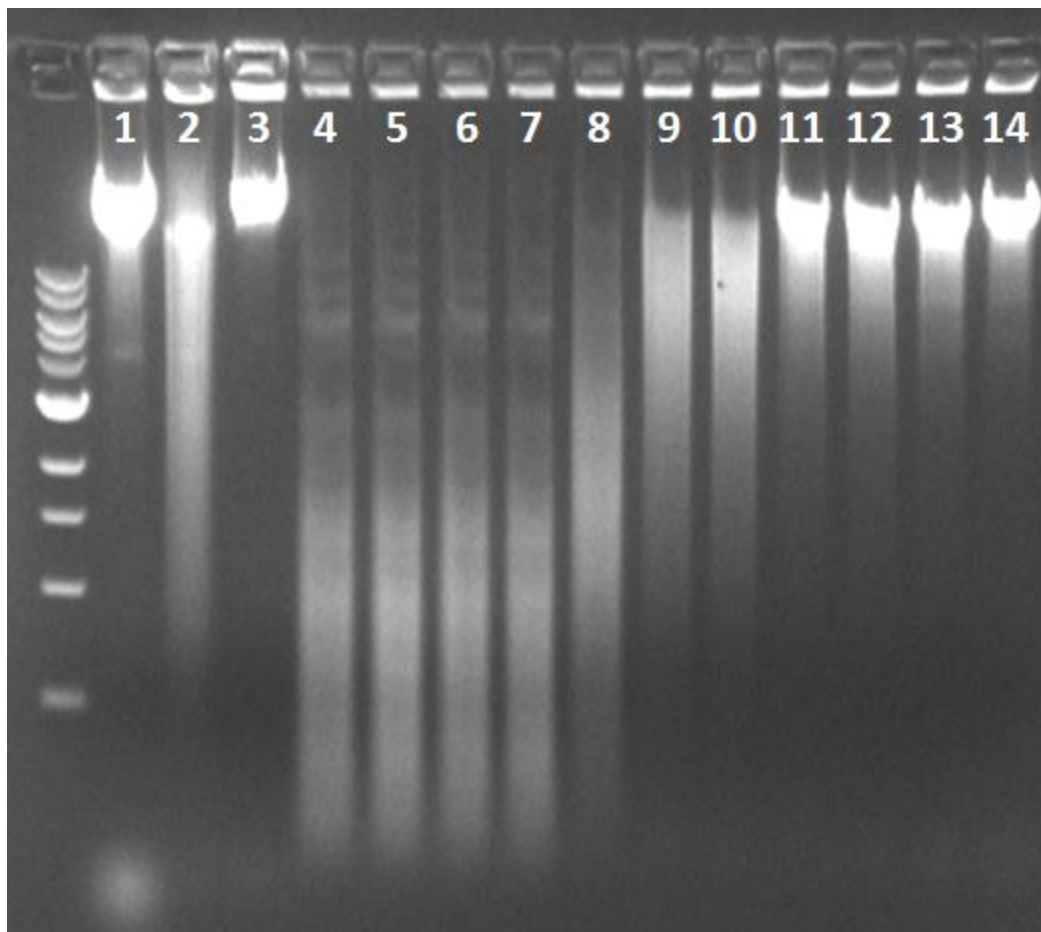


Figure 3.3: Partial digests of *S. cerevisiae* gDNA. The DNA was separated and visualized by gel electrophoresis in a 1 % agarose gel. The first lane to the left 1 kb ladder (bands corresponds to 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 8.0, and 10.0 kb) from New England Biolabs Inc.. Lane 1: gDNA preparation, lane 2: gDNA preparation incubated at 37 °C for 0.5 h, lane 3: gDNA mixed with BSA and buffer 4 followed by incubation in 37 °C for 0.5 h, and lane 4-14: 1:1 dilution series of Sau3AI with constant concentration of gDNA after incubation at 37 °C for 0.5 h.

The three digestion reactions exhibiting the desired size distribution were pooled and size-selected on a 0.75 % agarose gel, and the fragments in the size ranging from 1.5 to 10 kb were recovered. The recovered fragments were treated with Klenow (-exo) fragment and ligated into the pRS414TDH3 plasmid which prior to the ligation was digested with SalI-HF and treated with Klenow (-exo). The ligation mix was concentrated approximately 10-fold by precipitation and used to transform electro-competent *E. coli* cells (ElectroMAX™ DH10B™ T1 Phage-Resistant Competent Cells, $>1 \cdot 10^{10}$ CFU/ μ g DNA). The resultant transformants were plated on 10 large LB+amp plates and incubated overnight at 37 °C. It was estimated that each plate contained at least 10,000 colony forming units (CFU), and the total size of the working subset of the library is

hereby estimated to be approximately 100,000 CFU. Since only one tenth of the ligation mix was transformed into *E. coli*, the total library size is around $1 \cdot 10^6$ CFU. Sixteen colonies were picked and analyzed by colony PCR to determine if they had inserts. The result of the colony PCR can be seen in Figure 3.4. A band of around 600 bp indicates that the plasmid harbored by the tested cells is empty, whereas a larger band indicates that a DNA fragment has been inserted in the plasmid. From Figure 3.4 it can be seen that 13 out of 16 tested colonies contained an insert, corresponding to an insert percentage of 81.25 %.

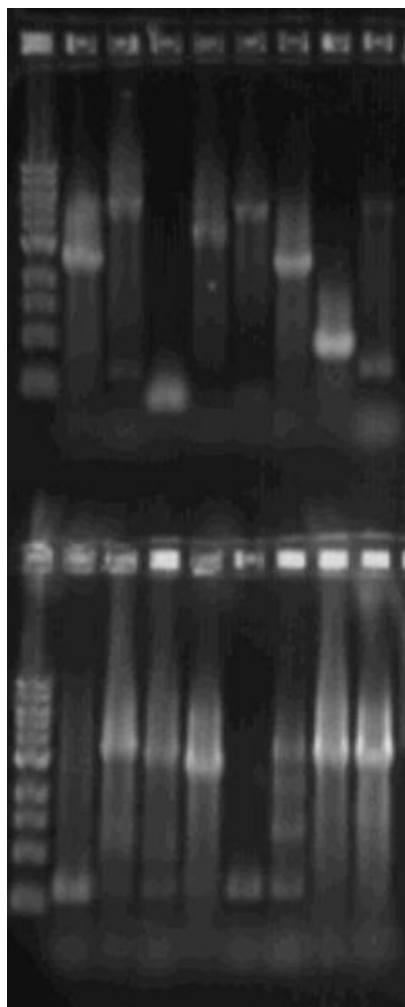


Figure 3.4: Visualization of colony PCR. Colonies were tested by PCR to determine the percentage of colonies harboring a plasmid which have had a gDNA fragment inserted. The primers were designed to anneal approximately 300 bp on each side of the insertion site. Bands at sizes larger than 600 bp indicate that a gDNA fragment has been inserted, whereas a band at around 600 bp indicates that no fragments have been inserted. The PCR was separated and visualized on a 1 % agarose gel. The first lane to the left in each tier is the 1 kb ladder (bands corresponds to 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 8.0, and 10.0 kb) from New England Biolabs Inc., and the remaining 16 lanes each represent an individual colony tested by colony PCR.

Some of the lanes in Figure 3.4 can be seen to have multiple bands which can be caused by two things, either by erroneous primer annealing or by the picking of multiple colonies into the colony PCR reaction. The latter is the more probable cause since 10,000 colonies on a single agar plate, even though they are evenly distributed, only allows for a limited area of growth for each colony and thereby increases the probability of joined-colonies forming from 2 or more transformants. Thus the insert percentage of 81.25 % is only approximating the actual insert percentage, but since this is always the case when a subset of a population is analyzed in order to characterize the entire population, this was not investigated more thoroughly. For further calculations the insert percentage will be rounded to 80 %. The colonies from the 10 plates were scraped off and purified, resulting in 0.5 µg/µl library pooled in a total volume of 0.5 ml.

3.4.3 Library quality considerations

High molecular weight DNA is prone to physical shearing which means that the purified DNA has to be handled very carefully. If the DNA is physically sheared rather than digested, the fragments can act as competitive inhibitors of the subsequent ligation reaction, so to ensure a favorable ratio between sheared and digested DNA, it is important that the size of the purified gDNA is at least 3- to 4- fold larger than the average size of the insert [Wu et al., 2004]. From Figure 3.4 the average insert size of our library can be estimated to be around 2 kb, and with an average size of purified gDNA > 10 kb (cf. Figure 3.3) the ratio requirement is met.

When fragmenting gDNA with restriction endonuclease, the distribution will be biased due to the uneven distribution of cleavage sites throughout the genome. This bias can be lessened by using endonucleases with short recognition sequences [Wu et al., 2004]. We applied the *Sau3AI* which has a 4 bp recognition sequence. In a completely random DNA sequence, it would cleave every 256 bp; however, restriction endonucleases with 4 bp recognition sequences would cleave within most ORFs if they were to fully digest the gDNA, which is the main reason for limiting the digestion time to ensure a partial digest. Furthermore, it is important to apply multiple concentrations of the endonuclease to avoid bias at individual cleavage sites and to obtain a more even concentration distribution over fragment sizes [Jauert et al., 2005].

To take GC-content of the gDNA into consideration, different endonucleases with 4 bp recognition sequences could be applied. Methods for completely avoiding the bias caused by cleavage site distribution such as random fragmentation of the gDNA by physical shearing or DNase I digestion have been developed [Ramer et al., 1992; Santangelo et al., 1986]. The drawback of these methods is that they generate random ends which have to be blunt-ended before they can be ligated into a plasmid, and thereby the ligation efficiency is decreased compared to 4 bp overhangs. It is possible to create 4 bp overhangs on the blunt-ended fragments by ligating adapters to each end of the fragments. However, this would require an additional ligation reaction, hereby lowering the overall yield. We chose to use the restriction endonuclease approach since we were more concerned with having a high ligation efficiency combined with

the subsequent possibilities that it creates for avoiding plasmid self-ligation, having plasmids with multiple inserts and DNA self-ligation.

The standard way of preventing plasmids from self-ligating is to treat the digested plasmid with phosphatase [Rodrigues et al. 2001]. This does only prevent the plasmid from self-ligating and thus does not prevent multiple inserts or fragment self-ligation. Furthermore, the efficiency of the phosphatase is crucial to obtaining a good library since linearized plasmids with only one of the two phosphate groups will be able to self-ligate albeit with a gap. Our initial library construction was performed with phosphatase-mediated dephosphorylation, and our best library constructed in this manner had an insert percentage around 20 % (result not shown), which is unacceptably low. To circumvent all of the three issues, the large Klenow fragment engineered to have no exonuclease activity can be used to partly fill in the 4 bp overhangs. This results in 2 bp overhangs which are unable to mediate any self-ligation and integration of multiple DNA fragments. However, this will lower the ligation efficiency but not to the same level as for the blunt-end case. For the case of plasmid self-ligation, which is the biggest problem of the three since it results in transformants harboring empty plasmids, the phosphatase treatment can create four possible outcomes: no dephosphorylation, dephosphorylation of either of the two ends of the plasmid, or dephosphorylation of both ends. Three of these four possibilities will be able to self-ligate, whereas the treatment with Klenow fragment will prevent plasmid self-ligation if a single nucleotide is filled in at either of the two ends. Actually the only plasmids capable of self-ligating would be linearized plasmids completely unaffected by the Klenow treatment, so if enzyme activity is disregarded, these considerations illustrate why Klenow treatment would be more efficient in preventing plasmid self-ligation, which is also supported by our results.

The average ORF size in *S. cerevisiae* was found to be 1,385 bp [Hurowitz and Brown, 2003] which is less than the average fragment size found for our gDNA library. Furthermore the probability of a given DNA sequence being present in a gDNA library can be estimated from the following formula:

$$p = 1 - e^{-\left(\frac{N \cdot I}{G}\right)}$$

where N is the total number of clones containing an insert, I is the average insert size and G is the target genome size (the combined exponent is the calculation of genome equivalents). The formula is based on a Poisson distribution [Moore, 1998] and shows that a greater number of clones obtained and/or a longer average insert size results in a greater probability of a given DNA sequence being present in the library. For the working subset of the library this probability was calculated as follows:

$$p = 1 - e^{-\left(\frac{8 \cdot 10^4 \cdot 2 \cdot 10^3}{12.5 \cdot 10^6}\right)} = 0.999997 \approx 1$$

In order for the potential ORFs inserted in the expression vector to be deregulated, they not only need to be present but also to be inserted in the correct direction according to the promoter (in this case the *TDH3* promoter), and the digestion site upstream of the start codon has to be in close enough proximity to the start codon so the native regulatory elements which could affect transcription are replaced with the ones present in the plasmid-based promoter sequence. Ramer et al. (1992) constructed a gDNA library using the inducible *GALI* promoter. To test their library, they plated it on selective media with galactose and auxotrophic selection for histidine prototrophic transformants. They obtained 42 clones which were subsequently plated on similar medium with glucose as the carbon source instead of galactose to limit transcription initiation from the *GALI* promoter. Of the 42 mutants, 4 were unable to grow on histidine-deficient medium thereby proving that these had indeed been deregulated. They fragmented their gDNA by physical shearing which of course means that there is an equal probability for each ORF to be deregulated. This example shows that this strategy of constructing overexpression libraries is feasible, be it with an inducible, repressible, constitutive or any other type of promoter. We applied random digestion, and there will therefore be some ORFs that we will be unable to deregulate due to the absence of a *Sau3AI* cleavage site between their start codons and their regulatory element(s). We applied a constitutive promoter and were hereby unable to perform a similar test since it is very difficult to set up an auxotrophic selection pressure which can discriminate between the native and the constitutive promoter. However, based on the results from Ramer et al. (1992) we are confident that we will be able to obtain deregulated versions of most of the ORFs present in the library. Nevertheless there will always be some genetic elements which cannot be obtained from a gDNA library due to their lethality or their poor maintenance and/or propagation in a heterologous host [Wu et al., 2004].

cDNA libraries would in many cases be a better choice of library compared to gDNA libraries since only genetic elements encoding proteins would be obtained, and the elements can be inserted in the correct direction into the expression plasmid. However, not all organisms perform poly-adenylation as part of their post transcriptional modifications, making the purification of their mRNA and subsequent cDNA library construction troublesome. Many of the organisms that do not poly-adenylate their mRNA also do not have introns. In these cases gDNA libraries would be the library type of choice, whereas for organisms with a high percentage of genes containing introns, cDNA libraries would be the best solution. In our case the donor organism and host organism are the same, thereby negating the need to construct a cDNA library. In the end, the best suited library type primarily depends on the organism from where the genetic element is sought.

3.4.4 Library screening and strain testing

The constructed library was transformed into the lycopene producing *S. cerevisiae* strain SimC002. Of the resultant transformants 200 colonies were picked based on their individual red color intensity by visual inspection and transferred onto new SC-TRP-URA plates. After incubation at 30 °C the new colonies, each corresponding to one of the initial 200, were screened a second time by visual inspection. Twenty seven colonies were selected for further testing to identify the ones capable of producing the most lycopene in liquid medium. Each colony was inoculated into 5 ml SC-TRP-URA medium and tested in technical triplicate, as was the SimC008 control strain. The mean values for the technical triplicates can be seen in Figure 3.5 where each mean has been normalized with respect to the SimC008 mean.

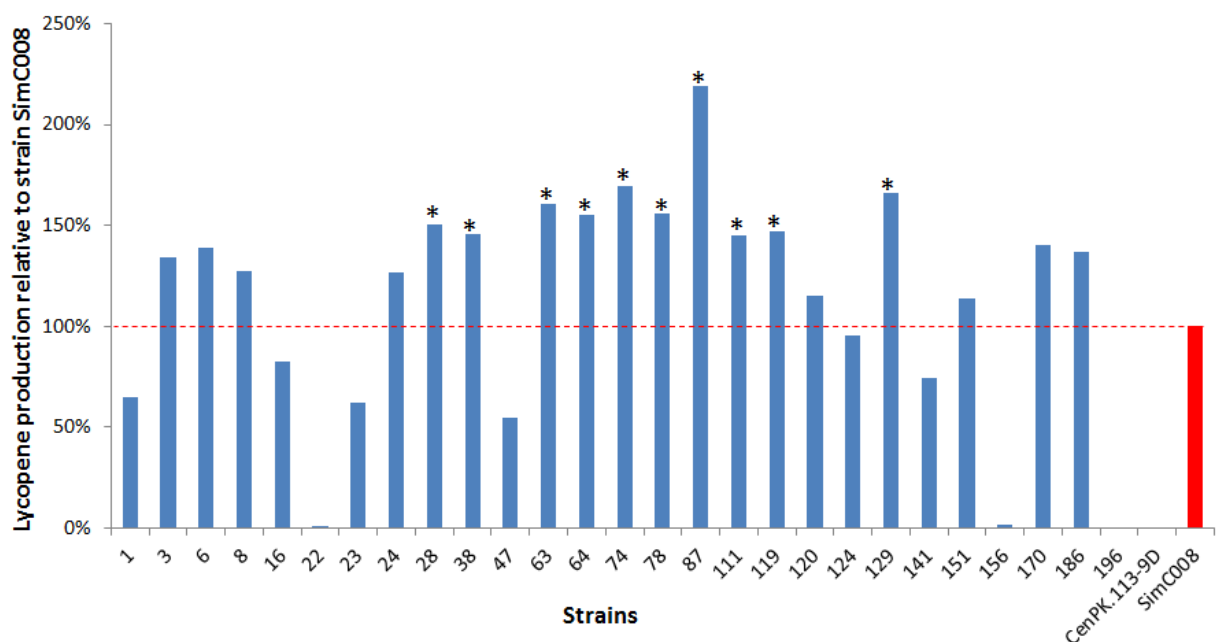


Figure 3.5: Relative isoprenoid production level. 27 of the 200 isolated mutants were tested for their ability to produce lycopene in 5 ml cultures together with the non-carotenogenic strain CEN.PK.113-9D and the positive control SimC008. The lycopene level were quantified in technical triplicate as a biomass yield (mg/l/OD₆₀₀) and normalized to the value of the positive control strain. The ten mutants with the highest biomass yield of lycopene are marked with an asterisk.

Eighteen of the 27 tested strains were shown to have higher lycopene producing capabilities compared to the SimC008 strain. Of these 18 strains the 10 best were selected for further evaluation (marked with an asterisk in Figure 3.5). Plasmids from the 10 strains were isolated and used to transform SimC002. From the resultant colonies three colonies were picked

randomly from each transformation and tested in liquid SC-TRP-URA medium for their lycopene production levels. The results can be seen in Figure 3.6 (blue columns).

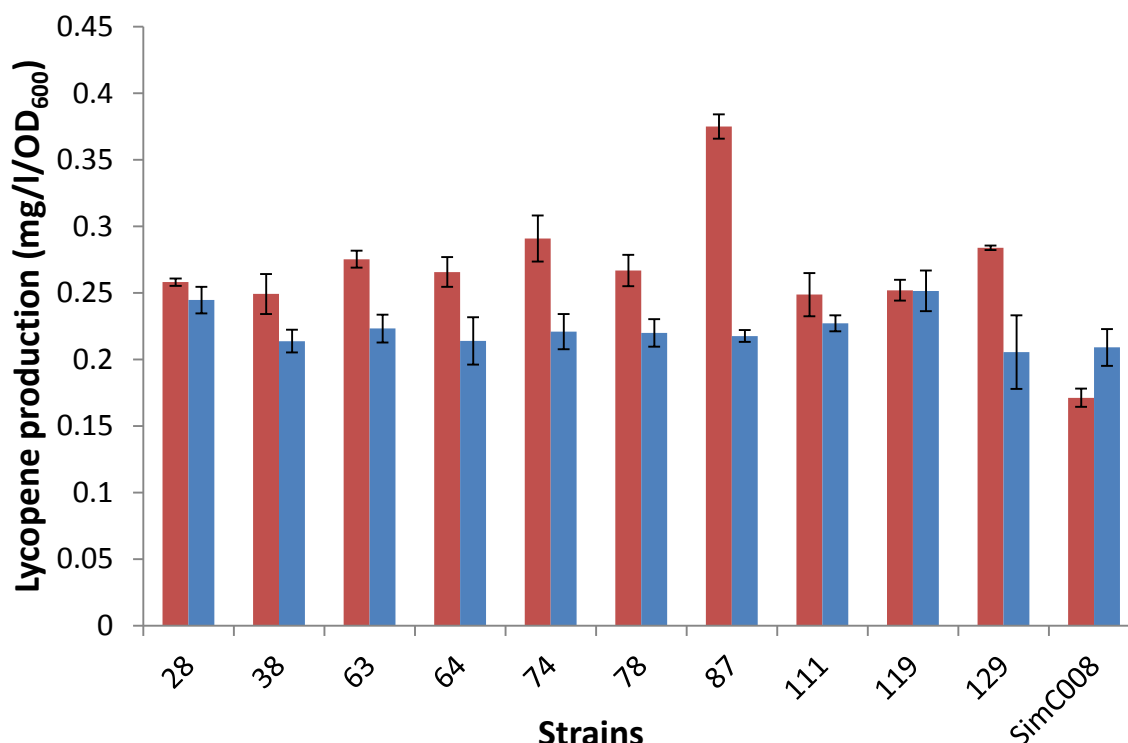


Figure 3.6: Lycopene biomass yield determination. The red columns are the initial non-normalized values determined for the ten best mutants shown in Figure 3.5, and the blue columns are the values determined for 3 randomly picked colonies of the retransformation of the clean background strain SimC002 with the plasmids isolated from the aforementioned ten mutants. The experiment was performed in 5 ml cultures and compared to the positive control strain SimC008. The error bars represent the standard deviation calculated for n=3 (technical triplicates for the initial values and biological triplicates for the reconstitution experiment).

For comparison the values for the initial test can be seen as the red columns in Figure 3.6. The only plasmid capable of re-constituting the lycopene overproducing phenotype is the plasmid isolated from colony 119 of the original transformation. Furthermore, it can be seen from the two tests that for the control strain there is a difference in the lycopene production level which is outside the limits delineated by the standard deviations. This could probably be caused by small differences in the media composition since a new batch was prepared for each of the two experiments, and it emphasizes how important it is to always include the control strain to identify the base lycopene production level for each individual batch of medium. The plasmid isolated from colony 119 was sequenced but found not to contain an open reading frame (results not

shown). Since the initial increase in lycopene production level could not be attributed to the genetic element harbored within the individual plasmids, the cause remains unidentified but could be caused by mutations or other genetic alterations that the individual cells had undergone during the transformation. These results demonstrate that even though mutants with increased lycopene production levels are isolated, there is no guarantee that the increase is caused by the transformed plasmid. In order to increase the probability of isolating such plasmids, the number of mutants analyzed in liquid cultures has to be increased.

Since one can only perform a given number of lycopene extractions in a given time frame, there will always be a tradeoff between how many individual transformants one can test and how many replicates it will be possible to include. From Figure 3.6 it can be seen that the standard deviations are relatively small, indicating that it could be possible to lower the number of replicates and still retain a certain degree of confidence in the results. This means that a higher number of transformants can be analyzed in the same time, and the probability of identifying lycopene overproducing mutants caused by the genetic element harbored within the transformed plasmid thereby could be increased as well.

Another possibility to increase the probability of identifying mutants with increased lycopene production could be to improve the screening method. Visual inspection of large numbers of colonies is very labor-intensive and also biased due to the individual color perception of the person performing the screen. If a high-throughput screening method independent of personal color perception could be set up, it would greatly aid in this endeavor.

3.4.5 Alternative screens

The increment in colony coloration as a function of carotenoid produced by the individual yeast colonies changes only slightly as carotenoid concentration increases, especially at higher carotenoid concentrations [Lewis et al., 1990]. This saturation effect can be mitigated by the addition of carotenoid biosynthetic inhibitors to the media. On this basis the carotenoid biosynthetic inhibitors diphenylamine [Chumpolkulwong et al., 1997] and β -ionone [Lewis et al., 1990] have been applied in chemical mutagenesis studies with subsequent screens for astaxanthin overproducers by visual inspection of *Xanthophyllomyces dendrorhous* mutants. The main problems with applying these inhibitors to identify mutants with increased isoprenoid production capabilities are that the inhibitors are specific to the carotenoid biosynthetic pathway, which would mean that most positive mutants isolated would probably be due to a mutation within the genes encoding this pathway, and that the inhibitors inhibit enzymes downstream of the lycopene synthase, which means that the heterologous pathway would have to be extended for them to be applied. Such genetic targets would of course be beneficial if the aim were to optimize *S. cerevisiae* for production of xanthophylls, but since the aim is to create an isoprenoid production platform with increased flux towards IPP, they would be useless. Other inhibitors such as statins have been shown to inhibit enzymes upstream of IPP and could therefore be used in the search for a general production host. However, this would mean that the search mainly

would aim at obtaining mutations within the gene encoding the protein targeted by the inhibitor. This “one step at a time”-approach would be very laborious, and it would therefore be preferential if the screen could identify mutations within multiple target at once. Subsequently, the identified targets could be engineered in a combinatorial approach to identify synergistic effects and by this mean identify the best candidate strain for a *S. cerevisiae*-based isoprenoid production platform.

The throughput of the screen could also be improved by applying a non-subjective standardized screening method capable of analyzing multiple mutants in shorter time than the manual inspection of agar plates. Digital Image Spectroscopy (DIS) has been demonstrated as a medium throughput method able to acquire and analyze absorption spectra in the region from 400-900 nm from up to 500 individual colonies [Arkin et al., 1990]. This method enables the researcher to analyze microbial populations for any property resulting in a change in absorption spectra such as the production of a pigment [Yang, 1994]. However, this method is, as mentioned above, only medium throughput since the colonies still have to be manually picked for further analysis. If this method were combined with a colony picker or another automatic method for separating the positive mutants from the remainder of the population, it would increase the throughput drastically, but a setup of this type must be considered very specialized; to our knowledge there have not been published any scientific paper for a method or setup like this. DIS could be combined with the inhibitor approach.

Fluorescence-activated cell sorting (FACS) would be the obvious choice if a very high-throughput screen is required. By FACS each individual cell can be analyzed and sorted at very high rates circumventing the need to plate the entire library of transformants to obtain colonies that can be analyzed. However, this approach relies on fluorescence, so a fluorescent surrogate isoprenoid metabolite needs to be produced heterologously as a reporter metabolite. The xanthophyll molecule astaxanthin has been found to be fluorescent and has been applied in a screen for *X. dendrorhous* mutants exhibiting increased astaxanthin production. The reported FACS methods can at best be considered as an enrichment step, which when applied to a library screen, has to be followed by a manual screen by visual inspection [An et al., 1991 and Ukibe et al., 2008]. This enrichment step will lower the number of colonies that have to be inspected and thereby also the work load, but in order for it to be applied in *S. cerevisiae*, the lycopene pathway has to be extended to convert lycopene into astaxanthin.

It is difficult to speculate which of the aforementioned methods for improving the screening effort is the most robust, but from a throughput point of view the FACS method seems to be the best choice, followed by the DIS method. If this is the case, in practice remains to be demonstrated.

3.5 Acknowledgements

The authors would like to acknowledge Dr. Hang Zhou for providing his help and guidance for the construction of the gDNA library.

3.6 References

Adrio JL, Demain AL: **Genetic improvement of processes yielding microbial products.** FEMS Microbiol. Rev. 2006, 20:187-214.

Ajikumar PK, Tyo K, Carlsen S, Mucha O, Phon TH, Stephanopoulos G: **Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms.** Mol. Pharmaceutics. 2008, 5(2):167–190.

Alper H, Jin YS, Moxley JF, Stephanopoulos G: **Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*.** Metab. Eng. 2005(a), 7:155-164.

Alper H, Miyaoku K, Stephanopoulos G: **Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets.** Nat. Biotechnol. 2005(b), 23(5):612-616.

Alper H, Stephanopoulos G: **Global transcription machinery engineering: a new approach for improving cellular phenotype.** Metab. Eng. 2007, 9:258-267.

An GH, Bielich J, Auerbach R, Johnson EA: **Isolation and characterization of carotenoid hyperproducing mutants of yeast by flow cytometry and cell sorting.** Biotechnology (N Y) 1991, 9:70-73.

Arkin AP, Goldman ER, Robles SJ, Goddard CA, Coleman WJ, Yang MM, Youvan DC: **Applications of imaging spectroscopy in molecular biology. II. Colony screening based on absorption spectra.** Biotechnology (N Y) 1990, 8: 746-749.

Bailey JE, Sburlati A, Hatzimanikatis V, Lee K, Renner WA, Tsai PS: **Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes.** Biotechnol. Bioeng. 1996, 52:109–121.

Britton G, Liaaen-Jensen S, Pfander H: **Carotenoids handbook**. Birkhäuser Verlag, Basel, Switzerland 2004.

Chemler JA, Koffas MAG: **Metabolic engineering for plant natural product biosynthesis in microbes**. Curr. Opin. Biotech. 2008, 19:597-605.

Chumpolkulwong N, Kakizono T, Nagai S, Nishio N: **Increased astaxanthin production by *Phaffia rhodozyma* mutants isolated as resistant to diphenylamine**. J. Ferment. Bioeng. 1997, 83(5):429-434.

Gallagher CE, Cervantes-Cervantes M, Wurtzel ET: **Surrogate biochemistry: use of *Escherichia coli* to identify plant cDNAs that impact metabolic engineering of carotenoid accumulation**. Appl. Microbiol. Biotechnol. 2003, 60:713-719.

Gietz RD, Woods RA: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method**. Method. Enzymol. 2002, 350:87-96.

Gray MR, Colot HV, Guarente L, Rosbash M: **Open reading frame cloning: identification, cloning, and expression of open reading frame DNA**. Proc. Natl. Acad. Sci.US. 1982, 79:6598-6602.

Hurowitz EH, Brown PO: **Genome-wide analysis of mRNA lengths in *Saccharomyces cerevisiae***. Genome Biol. 2004, 5(1):R2.

Jauert PA, Jensen LE, Kirkpatrick DT: **A novel yeast genomic DNA library on a geneticin-resistance vector**. Yeast. 2005, 22:653-657.

Jin YS, Alper H, Yang YT, Stephanopoulos G: **Improvement of xylose uptake and ethanol production in recombinant *Saccharomyces cerevisiae* through an inverse metabolic engineering approach**. Appl. Environ Microb. 2005, 71(12):8249-8256.

Jin YS, Stephanopoulos G: **Multidimensional gene target search for improving lycopene biosynthesis in *Escherichia coli***. Metab. Eng. 2007, 9:337-347.

Kang MJ, Lee YM, Yoon SH, Kim JH, Ock SW, Jung KH, Shin YC, Keasling JD, Kim SW: **Identification of genes affecting lycopene accumulation in *Escherichia coli* using a shot-gun method**. Biotechnol. Bioeng. 2005, 91(5):636-642.

Kirby J, Keasling JD: **Biosynthesis of plant isoprenoids: perspectives for microbial engineering**. Annu. Rev. Plant Biol. 2009, 60:335-355.

Lee SY, Lim HU, Park JH, Park JM, Kim TY: **Metabolic engineering of microorganisms: general strategies and drug production**. Drug Discov. Today. 2009, 14: 78-88.

Leonard E, Ajikumar PK, Thayer K, Xiao WH, Mo JD, Tidor B, Stephanopoulos G, Prather KLJ: **Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control**. Proc. Natl. Acad. Sci. US. 2010, 107(31):13654-13659.

Lewis MJ, Ragot N, Berlant MC, Miranda M: **Selection of astaxanthin-overproducing mutants of *Phaffia rhodozyma* with β -ionone**. Appl. Environ Microb. 1990, 56(9):2944-2945.

Lin D, Qu L-J, Gu H, Chen Z: **A 3.1-kb genomic fragment of *Bacillus subtilis* encodes the protein inhibiting growth of *Xanthomonas oryzae* pv. *Oryzae***. J. Appl. Microbiol. 2001, 91:1044-1050.

Mayer MP: **A new set of useful cloning and expression vectors derived from pBluescript**. Gene. 1995, 163:41-46.

Moore DD: 1998. **Construction of recombinant DNA libraries.** In Current Protocols in Molecular Biology, Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. (eds). John Wiley & Sons Inc.

Mumberg D, Müller R, Funk M: **Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds.** Gene. 1995, 156(1): 119-122.

Munterdam R, Melillo E, Ryden A, Kayser O: **Perspectives and limits of engineering the isoprenoid metabolism in heterologous hosts.** Appl. Microbiol. Biotechnol. 2009, 84:1003-1019.

Porro D, Gasser B, Fossati T, Maurer M, Branduardi P, Sauer M, Mattanovich D: **Production of recombinant proteins and metabolites in yeasts. When are these systems better than bacterial production systems?** Appl. Microbiol. Biotechnol. 2011, 89:939–948.

Ramer SW, Elledge SJ, Davis RW: **Dominant genetics using a yeast genomic library under the control of a strong inducible promoter.** Proc. Natl. Acad. Sci. US. 1992, 89: 11589-11593.

Rodrigues F, Zeeman A-M, Alves C, Sousa MJ, Steensma HY, Corte-Real M, Leao, C: **Construction of a genomic library of the food spoilage yeast *Zygosaccharomyces bailii* and isolation of the β -isopropylmalate dehydrogenase gen (*ZbLEU2*).** FEMS Yeast Res. 2001, 1:67-71.

Rose MD, Broach JR: **Cloning genes by complementation in yeast.** Method. Enzymol. 1991, 194:195-230.

Ruby SW, Szostak JW, Murray AW: **Cloning regulated yeast genes from a pool of *lacZ* fusions.** Method. Enzymol. 1983, 101:253-269.

Santangelo GM, Tornow J, Moldave K: **Cloning of open reading frames and promoters from the *Saccharomyces cerevisiae* genome: construction of genomic libraries of random small fragments.** Gene. 1986, 46:181-186.

Sedkova N, Tao L, Rouviere PE, Cheng, Q: **Diversity of carotenoid synthesis gene clusters from environmental Enterobacteriaceae strains.** Appl. Environ Microb. 2005, 71(2):8141–8146.

Sheff MA, Thorn KS: **Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*.** Yeast. 2004, 21:661-670.

Sikorski RS, Hieter PA: **A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*.** Genetics. 1989, 122:19-27.

Tao L, Jackson RE, Rouvière PE, Cheng Q: **Isolation of chromosomal mutations that affect carotenoid production in *Escherichia coli*: mutations alter copy number of ColE1-type plasmids.** FEMS Microbiol. Lett. 2005, 243:227-233.

Ukibe K, Katsuragi T, Tani Y, Takagi H: **Efficient screening for astaxanthin-overproducing mutants of the yeast *Xanthophyllomyces dendrorhous* by flow cytometry.** FEMS Microbiol. Lett. 2008, 286:241:248.

Weinstock GM, Ap Rhys C, Berman ML, Hampar B, Jackson D, Silhavy TJ, Weisemann J, Zweig M: **Open reading frame expression vectors: a general method for antigen production in *Escherichia coli* using protein fusions to β -galactosidase.** Proc. Natl. Acad. Sci. US. 1983, 80: 4432-4436.

Williams DH, Stone MJ, Hauck PR, Rahman SK: **Why are secondary metabolites (natural products) biosynthesized?** J. Nat. Prod. 1989, 52(6):1189–1208.

Wu C, Xu Z, Zhang H-B: 2004. **DNA libraries**. Pages 385-425 in: Encyclopedia of Molecular Cell Biology and Molecular Medicine, 2nd ed., vol. 3. Meyers RA, eds. Wiley-VCH Verlag GmbH, Weinheim, Germany.

Yang MM: **Digital imaging spectroscopy of microbial colonies**. Am. Biotechnol. Lab. 1994, 12(6):18-20.

Zhang YHP, Himmel ME, Mielenz, JR: **Outlook for cellulose improvement: screening and selection strategies**. Biotechnol. Adv. 2006, 24:452-481.

Chapter 4 – Carotenoid biosynthesis in *Saccharomyces cerevisiae*: potential application of fluorescence-activated cell sorting for identification of isoprenoid hyper-producing mutants

4.1 Abstract

Multiple techniques have been developed for generation of diversity within microbial populations and have greatly aided in the identification of novel genetic targets for obtaining microorganisms with desired phenotypes. As a result, very large mutant libraries can readily be created, but when searching these libraries for traits which cannot be actively selected for, the existence of an efficient screen is imperative for success. The screening method has to have sufficiently high throughput to be able to process the large number of mutants, which can readily be created. One of the screening methods with the highest throughput developed to date is the fluorescence-activated cell sorting (FACS). This method can be used to sort out microbial cells overproducing a fluorescent reporter molecule. We here demonstrate the engineering of *Saccharomyces cerevisiae* for the production of lycopene, β -carotene, and astaxanthin. The three constructed strains produced 1.05, 0.815, and 1.06 mg/gDCW lycopene, β -carotene, and astaxanthin, respectively. The obtained astaxanthin production is 9-fold higher than what has previously been achieved in *S. cerevisiae*. β -carotene and astaxanthin have previously been demonstrated to exhibit fluorescence when excited at 488 nm. The constructed strains served as the basis for testing if any of the three carotenoids as part of a sub-group of the isoprenoids can be applied as surrogate molecules for screening microbial libraries for increased isoprenoid production using FACS. The results show that the change in *in vivo* fluorescence due to production of carotenoids was too weak to enable us to efficiently apply these compounds as surrogate molecules in the endeavor of screening populations of *S. cerevisiae* for isoprenoid overproducers by FACS.

4.2 Introduction

Within the realm of metabolic engineering, there are two main approaches to engineering organisms for obtaining mutants exhibiting desired phenotypes, namely rational and combinatorial metabolic engineering. The rational approach relies on prior knowledge of the system to identify key targets which when perturbed will result in the desired phenotype improvements. Whereas, the combinatorial approach does not rely on prior knowledge but instead relies on the large number of techniques, from random mutagenesis to molecular biological methods such as gDNA, cDNA, and transposon libraries for generating diversity which can subsequently be screened to identify mutants with desired phenotypes [Tyo et al., 2006]. However, biological systems are highly regulated and intertwined creating an enormous search space which needs to be efficiently probed in order to improve the probability of identifying mutants with beneficial traits, which in turn means that mutant libraries need to be sufficiently large and diverse. The large number of mutants that needs to be screened in order for this approach to be effective necessitates the existence of a high-throughput screen.

One of the most effective screening methods developed is fluorescence-activated cell sorting (FACS), which allows for the characterization of individual cells by flow cytometry in conjunction with cell sorting, isolating the cells identified by the flow cytometer as exhibiting a desired trait. FACS was originally used in the study of mammalian cells but has been adapted to enable the characterization and sorting of microbial populations [Katsuragi and Tani, 2000]. The main objectives for FACS when applied for identification of cells with over-producing phenotypes are: reduce the required workload, reduce the required time, and identify the cells with the highest production rate [Mattanovich and Borth, 2006], which is ideal when screening large mutant libraries in the search of improved production phenotypes. A prerequisite for the application of FACS is that the target molecule fluoresces upon excitation or that a specific fluorescent probe for the molecule of interest exists. For a review on FACS see [Ibrahim and van den Engh, 2007].

The obvious target for screening for *S. cerevisiae* isoprenoid hyper-producers would be the two universal isoprenoid precursors, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). However, neither of these two molecules exhibits fluorescence, nor has any fluorescent probe for the two been identified. To bypass this problem a fluorescent surrogate isoprenoid molecule could be exploited. It has previously been shown that *Dunalella salina* and *Xanthophyllomyces dendrorhous* (formerly *Phaffia rhodozyma*) mutant hyper-producers of β -carotene and astaxanthin, respectively, could be isolated using FACS [Nonomura and Coder, 1988; An et al., 1991; Ukibe et al., 2008]. Besides FACS, laser confocal microscopy has been used to correlate the cellular auto-fluorescence with the carotenoid content of single *X. dendrorhous* cells [An et al., 2000]. All the aforementioned studies on carotenoid fluorescence applied a 488 nm argon ion laser.

Both β -carotene and astaxanthin belong to the group of secondary metabolites known as carotenoids, with astaxanthin belonging to the carotenoid sub-group known as xanthophylls, which encompasses oxygen decorated carotenoids. The main feature of carotenoids is their 9-11 conjugated C=C double bonds which creates a large π -electron system absorbing light within the gap of chlorophyll absorption making them ideal as accessory light-harvesting molecules in photosynthesis. This light absorption is also the cause for the carotenoids colorfulness and their application as pigments [Vershinin, 1999]. The absorption spectra of β -carotene and astaxanthin, together with their fluorescent excitation and emission spectra can be seen in Figure 4.1.

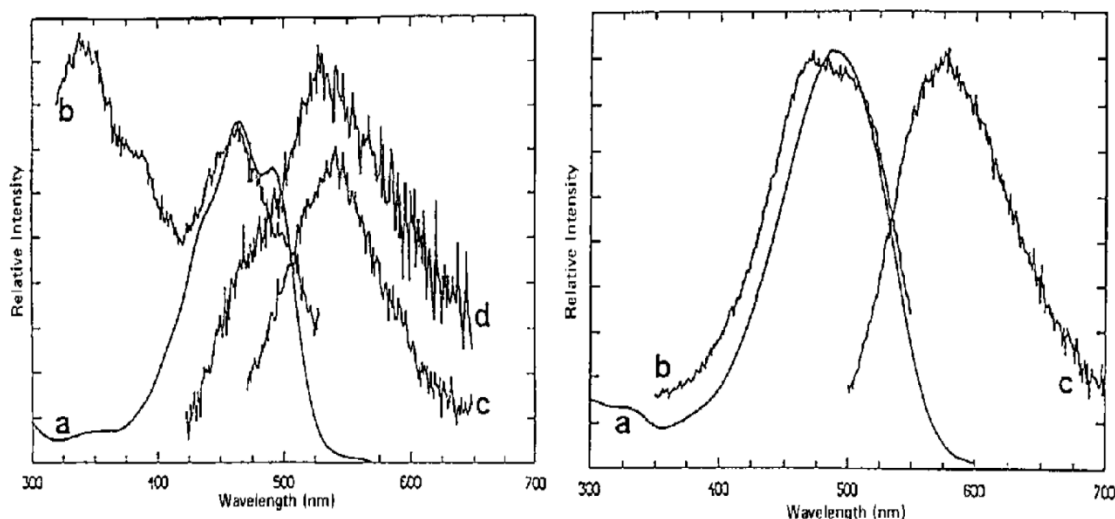


Figure 4.1: Absorption and fluorescence spectra of β -carotene (left panel) and astaxanthin (right panel) in chloroform at 25 °C. a: absorption spectra, b: fluorescence excitation spectra, c and d: emission spectra. The emission spectra for β -carotene were recorded at 430 nm (c) and 350 nm (d) excitation wavelengths; the excitation spectrum was recorded for an emission wavelength of 570 nm and normalized to the absorption maximum at 464 nm. The astaxanthin emission spectrum was recorded at 435 nm excitation wavelength; the excitation spectrum was recorded for an emission wavelength of 600 nm and normalized to the absorption maximum at 490 nm. The figure is adapted from [Jørgensen et al., 1992].

The structures of β -carotene and astaxanthin are very similar with the only differences being the hydroxyl moieties at the 3/3' positions and the keto moieties at the 4/4' positions. The structures of the molecules can be seen in Figure 4.2. The hydroxyl groups do not affect the optical properties of astaxanthin. However, the keto groups are conjugated with the polyene backbone and are responsible for the change in both the absorption and fluorescence spectra. The β -carotene absorption spectrum has the characteristic absorption maxima at 425, 450 and 475 nm whereas the astaxanthin absorption spectrum has a single absorption maximum at 460 nm.

Furthermore, the keto groups create a shift in the fluorescence spectra towards higher excitation and emission wavelengths. It should be noted that the excitation maxima coincide with the absorption maxima [Jørgensen et al., 1992; Visser et al., 2005]. The large number of double bonds allows for a great number of cis/trans isomers which have slightly different optical properties, but carotenoids in nature are predominately found in all-trans conformation, since this conformation is more thermodynamically favorable. The existence of two chiral centers at 3 and 3' positions give rise to three isomeric forms, namely the 3R/3'R, 3S/3'S and 3S/3'R. These optical isomeric forms do not exhibit differences in optical absorbance [Visser et al., 2005].

Biosynthesis of β -carotene and astaxanthin has been found to occur in several microorganisms, e.g. yeasts: *Sporobolomyces roseus* [Davoli and Weber, 2002], *X. dendrorhous* [Andrews et al., 1976]; green algae: *Haematococcus pluvialis* [Boussiba and Vonshak, 1991]; and marine bacteria: *Paracoccus* sp. strain N81106, *Paracoccus* sp. strain PC1 [Berry et al., 2003] and *Brevundimonas* sp. Strain SD212 [Nishida et al., 2005]. Since β -carotene and astaxanthin are both isoprenoids, they are both biosynthesized from the ubiquitous isoprenoid precursor IPP which is polymerized with three molecules of DMAPP to form geranylgeranyl pyrophosphate (GGPP). Two molecules of GGPP are then condensed head to head to form phytoene which is then desaturated to form lycopene. Subsequently each end of lycopene is cyclized to form β -carotene, which is finally decorated at the 3/3'-positions with hydroxyl moieties and at the 4/4'-positions with keto moieties to form astaxanthin. The course of the pathway until β -carotene is identical within prokaryotes and eukaryotes; however, the synthesis of astaxanthin from β -carotene proceeds differently in different organisms. In *X. dendrorhous* the reaction is catalyzed by a single enzyme, astaxanthin synthase encoded by the *crtS* gene. The *crtS* encodes a cytochrome P450 enzyme which in order to be functional has to be coupled to a reductase. The reductase corresponding to the *crtS* is encoded by the *crtR* gene. On the other hand, bacteria have been found to utilize two enzymes, namely 3,3'- β -carotene hydroxylase and 4,4'- β -carotene ketolase encoded by the *crtZ* and *crtW* genes, respectively [Nishida et al., 2005, Ukibe et al., 2009]. The bacterial biosynthetic pathway for β -carotene and astaxanthin can be seen in Figure 4.2.

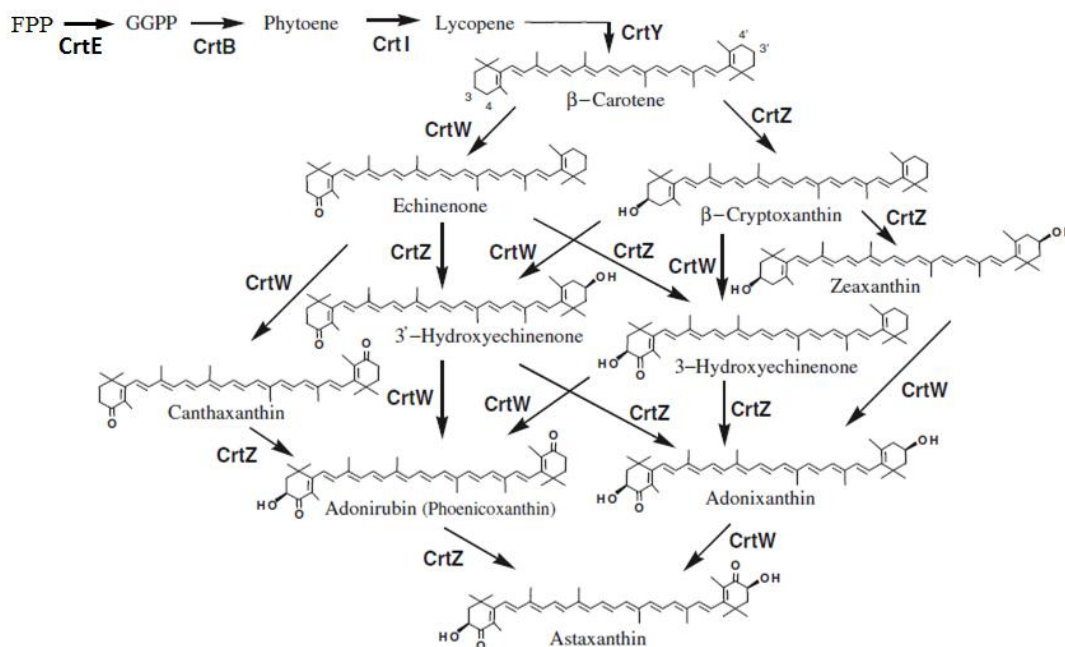


Figure 4.2: Carotenoid biosynthetic pathway to astaxanthin. FPP is converted to GGPP with one molecule of DMAPP by CrtE, followed by the condensation of two molecules of GGPP, catalyzed by the CrtB, to form phytoene. Subsequently, catalyzed by the CrtI, phytoene is desaturated to form lycopene followed by the cyclization by CrtY of each end of the lycopene molecule to form β -carotene. β -carotene undergoes a series of oxidative reactions to finally form astaxanthin. These reactions are catalyzed by CrtZ and CrtW. The figure is modified from [Choi et al., 2006].

Heterologous microbial production of β -carotene and astaxanthin has been achieved in *Escherichia coli*, *Saccharomyces cerevisiae* and *Candida utilis*. For reviews on heterologous microbial production of carotenoids see Misawa and Shimada (1998) and Das et al. (2007). For microbial β -carotene production in *S. cerevisiae* the *X. dendrorhous crtE*, *crtYB*, and *crtI* genes have been successful in yielding up to 5.9 mg/gDCW in an engineered background strain [Verwaal et al., 2007; Lange and Steinbüchel, 2011]. However, the bacterial *crtE*, *crtB*, *crtI*, and *crtY* genes have not been as effective in this endeavor. Yamano et al. (1994) were only able to obtain 103 μ g/gDCW β -carotene in *S. cerevisiae* with genes from *Pantoea ananatis* (formerly called *Erwinia uredovora*), but codon optimized versions of these genes have been applied in *C. utilis* with greater success yielding up to 0.4 mg/gDCW β -carotene [Miura et al., 1998]. In the same study it was demonstrated that the produced β -carotene could be converted to astaxanthin by co-expressing codon optimized versions of the *crtZ* and *crtW* genes from *Agrobacterium aurantiacum*. Besides this example, only one other example of astaxanthin production in yeast has been published in the literature, namely the study by Ukibe et al. (2009). They expressed the *crtE*, *crtYB*, *crtI*, *crtS* and *crtR* genes from *X. dendrorhous* in *S. cerevisiae* which resulted in the production of 120 μ g/gDCW astaxanthin. They furthermore tested the substitution of the *crtS*

and *crtR* genes with the *crtZ* gene from *Pantoea ananatis* and *crtW* from *Paracoccus* sp. NBRC101723. This substitution resulted in a production of 29 µg/gDCW astaxanthin.

We here present the construction of two *S. cerevisiae* strains for production of β-carotene and astaxanthin, respectively. This was done on the basis of a previously constructed plasmid which enables *S. cerevisiae* to produce lycopene. To extend the pathway to β-carotene and astaxanthin, we created two plasmids for chromosomal integration of codon optimized versions of the *crtY* gene and the *crtY*, *crtZ*, and *crtW* genes, respectively. The codon optimized genes are based on the corresponding genes from *Brevundimonas* sp. SD212. Use of bacterial genes was selected since heterologous expression of cytochrome P450 (required for the *X. dendrorhous* pathway) is inherently troublesome. The reason that the *Brevundimonas* sp. SD212 genes were chosen as the basis for the codon optimized genes is that the *crtZ* and *crtW* from this organism has been demonstrated to be more effective at producing astaxanthin compared to the corresponding genes from *Paracoccus* sp. PC1, *Paracoccus* sp. N81106 and *Pantoea ananatis* when heterologously expressed in *E. coli* [Choi et al., 2005; Choi et al., 2006]. Since Miura et al. (1998) have demonstrated that codon optimizing the carotenoid biosynthetic genes from *Pantoea ananatis* results in higher carotenoid production we decided to only test the codon optimized versions of the *Brevundimonas* sp. SD212 genes in *S. cerevisiae*. We chose to integrate the genes in the chromosome of *S. cerevisiae* since it has been demonstrated that chromosomal integration of the carotenoid biosynthetic genes improved the production of β-carotene [Verwaal et al., 2007]. The constructed strains finally form the basis for testing whether FACS can be utilized to discriminate between *S. cerevisiae* strains producing lycopene, β-carotene, or astaxanthin and a non-carotenogenic negative control strain.

4.3 Material and methods

4.3.1 Enzymes, chemicals and media

All enzymes applied for constructing plasmids were purchased from New England Biolabs Inc., Ipswich, MA. The lycopene and β -carotene standards were purchased from Sigma-Aldrich®, St. Louis, MO, and the astaxanthin standard was purchased from Enzo Life Sciences, Inc., Farmingdale, NY. The HPLC grade water and acetone used for the HPLC analysis were purchased from Sigma-Aldrich®. *S. cerevisiae* were grown on SC medium with the appropriate nutrient(s) omitted. SC medium was composed of 6.7 g/l Yeast Nitrogen Base (YNB) without amino acids, Sunrise Science Products, Inc., San Diego; CSM – amino acid dropout mixture (amount according to manufacturer's recommendation), Sunrise Science Products, Inc., and 20 g/l dextrose. *E. coli* cells were grown in BD Difco™ LB broth from BD Diagnostic Systems, Sparks, MD. For selection and maintenance of plasmids, ampicillin was added to a final concentration of 100 mg/l.

4.3.2 Plasmid and strain construction

The codon optimized genes encoding the CrtY (GenBank: BAD99407.1), CrtZ (GenBank: BAD99414.1), and CrtW (GenBank: BAD99406.1) enzymes were purchased from Bio Basic Inc., Markham, Ontario, Canada. The genes were synthesized to be flanked with SpeI and XhoI restriction sites, which were used to clone the genes into pSC011, pSC012, and pSC013 hereby creating the *TEF1p-crtY-ADH1t*, *PGK1p-crtY-ACT1t*, and *TDH3p-crtW-CYC1t* expression cassettes, respectively. The *TEF1p-crtY-ADH1t* expression cassette was amplified by PCR using Phusion® High-Fidelity DNA Polymerase and the following primers: 5'-AAGCTGGAGCTCATAGCTTCAAAATGTTTCTACTCCTTTTTTAC-3' and 5'-TATGACGCGGATCCGCGGCCGCTGACCTAGACCTAGGGAGCGACCTCATGCTATACCTG-3' (restriction sites are underlined). The amplified DNA fragment was treated with SacI and BamHI endonucleases and ligated into pRS405 which prior to the ligation also had been treated with SacI and BamHI endonucleases and gel purified. This resulted in the construction of pSCA101 which contains the *TEF1p-crtY-ADH1t* expression cassette followed by the AvrII and NotI restriction sites. The *PGK1p-crtY-ACT1t* expression cassette was amplified with the following primers: 5'-GACTTAGCTAGCAGACGCGAATTTTTCGAAGAAGTACC-3' and 5'-TATGACGCGGCCGCTGACCTAGACCTAGGCCGATATGATACACGGTCCAATG-3' (restriction sites are underlined). The amplified fragment was treated with NheI and NotI endonucleases and ligated into pSCA101 which prior to the ligation had been treated with AvrII and NotI endonucleases and gel purified. This resulted in the construction of pSCA102 which contains the *TEF1p-crtY-ADH1t* and *PGK1p-crtY-ACT1t* expression cassettes. The NheI endonuclease used to treat the insert generates 4 bp overhangs which are compatible with the 4 bp overhangs generated by the AvrII endonuclease used to treat the backbone; therefore the final construct will neither have the NheI nor the AvrII restriction sites. However the reverse primer

used to amplify the expression cassette contained a new *AvrII* restriction site, allowing this to be repeated and permitting iterative plasmid construction. The *TDH3p-crtW-CYC1t* expression cassettes were amplified using the following primers: 5'-GACTTAGCTAGCAGTTTATCATTATCAAT-3' and 5'-TATGACGCGGCCGCTGACCTAGACCTAGGAAATTGTAAACGTTAATATTTTGTTAAATTTCGCGTTAA-3' (restriction sites are underlined). The amplified fragment was inserted in pSCA102 to create pSCA103 the same way as the previous fragment was inserted into pSCA101. All ligations were performed overnight at 16 °C using T4 DNA ligase. The constructed plasmids can be seen in Table 4.1 and the plasmid maps in Figure 4.3.

All transformations of *S. cerevisiae* were done using the lithium acetate and polyethylene glycol method according to Gietz and Woods (2002). Auxotrophic selection of the transformants was done on SC medium with the appropriate nutrients(s) omitted. Verification of the constructed strains was done by extraction of gDNA using the Wizard® Genomic DNA Purification Kit, Promega, Madison, WI, followed by PCR amplification of all the integrated ORFs. The constructed strains can be seen in Table 4.1.

Table 4.1 - Strains and plasmids. The table shows the acquired and constructed, strains and plasmids applied in this study.

Strain	Genotype	Source
<i>Saccharomyces cerevisiae</i>		
CEN.PK113-7D	<i>MATa MAL2-8^c SUC2</i>	Dr. P. Kötter, Frankfurt, Germany.
CEN.PK102-3A	<i>MATa MAL2-8^c SUC2 leu2 ura3</i>	Dr. P. Kötter, Frankfurt, Germany.
SimC013	<i>MATa MAL2-8^c SUC2 leu2::pRS405(LEU2) ura3::pSC203(TEF1p-CrtE-ADH1t PGK1p-CrtB-ACT1t TDH3p-CrtI- CYC1t URA3)</i>	This study.
SimC014	<i>MATa MAL2-8^c SUC2 leu2::pSCA101(TEF1p-CrtY-ADH1t LEU2) ura3::pSC203(TEF1p-CrtE- ADH1t PGK1p-CrtB-ACT1t TDH3p- CrtI-CYC1t URA3)</i>	This study.
SimC015	<i>MATa, MAL2-8^c, SUC2, leu2::pSCA103(TEF1p-CrtY-ADH1t PGK1p-CrtZ-ACT1t TDH3p-CrtW- CYC1t LEU2) ura3::pSC203(TEF1p- CrtE-ADH1t PGK1p-CrtB-ACT1t TDH3p-CrtI-CYC1t URA3)</i>	This study.
<i>Escherichia coli</i>		
MAX Efficiency® DH5α™	F- φ80lacZΔM15 Δ(lacZYA-argF) U169 <i>recA1 endA1 hsdR17</i> (rk ⁻ , mk ⁺) <i>phoA supE44 λ⁻ thi⁻1 gyrA96 relA1</i>	Life Technologies, Grand Island, NY.
Plasmid name		Origin/reference
pSC203		Center for Microbial Biotechnology plasmid collection, DTU†.

Carotenoid biosynthesis in *Saccharomyces cerevisiae*: potential application of fluorescence-activated cell sorting for identification of isoprenoid hyper producing mutants

pRS405	[Sikorski and Hieter, 1989]
pSC011	Biotechnology plasmid collection, DTU†.
pSC012	Biotechnology plasmid collection, DTU†.
pSC013	Biotechnology plasmid collection, DTU†.
pSCA101	This study.
pSCA102	This study.
pSCA103	This study.

† Constructed by Simon Carlsen as part of his Ph.D. thesis work.

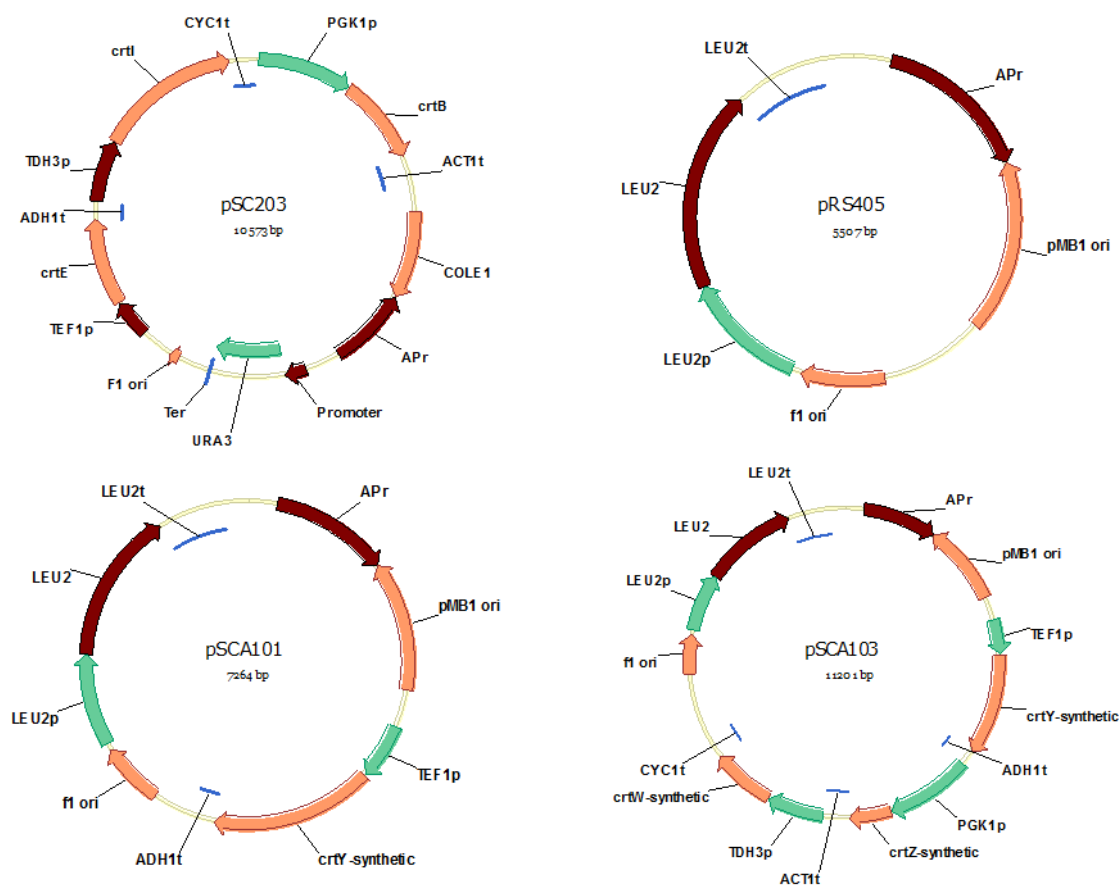


Figure 4.3: Plasmid maps for the plasmids utilized in this study.

4.3.3 Carotenoid extraction

The yeast strains (CEN.PK113-7D, SimC013, SimC014 and SimC015) were tested for their ability to produce carotenoids by growing them on 6 ml SC-LEU-URA in 14 ml Falcon 2059 tubes, BD Biosciences, Sparks, MD. The cells were inoculated at $OD_{600} = 0.1$ from pre-cultures and incubated at 30 °C with 250 rpm orbital shaking for 2 days. The OD_{600} was measured spectrophotometrically, followed by harvesting of the cells. For verification of the carotenoids produced by the cells, 5 ml of the cultures were harvested, whereas for quantification, 1 ml of the cultures was harvested. The cultures were harvested by centrifugation at 3000 g for 10 min, washed with 1 ml phosphate-buffered saline (Cellgro®, Mediatech, Inc., Manassas, VA) and transferred to a 2 ml FastPrep® tube (MP Biomedicals LLC, Solon, OH). The cells were spun down at 18,000 g for 1 min and the phosphate-buffered saline removed. Two hundred μ l acid washed glass beads (425-600 μ m, Sigma-Aldrich®) and 1 ml of acetone was added to each FastPrep® tube. The cells were disrupted using a Savant FastPrep® FP120 Cell Disrupter (MP Biomedicals LLC.), and the extracts were filtered using a 13 mm syringe filter (w/ 0.2 μ m PTFE filter, VWR, Radnor, PA). Finally 800 μ l of each of the extracts were transferred to HPLC vials for further analysis.

4.3.4 Carotenoid analysis and quantification by HPLC

The HPLC analysis of the carotenoids was performed according to the method described by Rentel et al. (1998). The method was performed on a Shimadzu UFLC-LC-20AD equipped with an Agilent Hypersil AA-ODS column (5 μ m, 2.1 x 200 mm) and appertaining guard column. Chromatograms were recorded using a SPD-M20A diode array detector, detecting at 474 nm together with scans obtained in the interval 350-550 nm. The chromatography was performed with a constant flow at 0.750 ml/min with the column at room temperature. Following injection of the sample the mobile phase was kept at 10 % water in acetone for 1 min, after which the acetone concentration was increased to 100 % using a linear gradient for 4 min. The concentration was kept at 100 % for 2 min, and subsequently decreased in 1 min to the initial concentration using a linear gradient followed by 2 min hold at this concentration before the next sample was injected. The method was validated using authentic standards of lycopene, β -carotene and astaxanthin which were dissolved in acetone and mixed in a single HPLC vial. A dilution series of the carotenoid standard mix were made in triplicates and analyzed using the method described above to obtain correlations between the individual areas under the peaks with their corresponding carotenoid concentrations enabling quantification of carotenoids.

4.3.5 Fluorescence microscopy

The yeast cells were inoculated in 5 ml SC-LEU-URA medium at $OD_{600} = 0.1$ and incubated overnight at 250 rpm orbital shaking. The cells from 1 ml of each culture was transferred to a micro centrifuge tube and harvested at 18,000 g for 1 min. The cells were washed with phosphate-buffered saline and re-suspended in 100 μ l phosphate-buffered saline. Five μ l of each cell suspension was transferred to a microscope slide and examined using the Nikon Eclipse TE200-S microscope, Nikon Inc., Melville, NY equipped with 100x objective and GFP band-pass filter (GFP-B: EX480/40, DM505, BA535/50). Fluorescent images were acquired using 10 sec exposure time and processed using the SPOT AdvanceTM software, SPOTTM, Imaging Solutions, a division of Diagnostic Instruments, Inc., Sterling Heights, MI.

4.3.6 Flow cytometry

The yeast cells were grown the same way as for the fluorescence microscopy. The cells were washed with phosphate-buffered saline and re-suspended in phosphate-buffered saline at an approximate concentration of $1 \cdot 10^6$ cells per ml. The cells were analyzed using the FACS caliber cytometer, BD Biosciences equipped with a 488 nm Argon-ion blue laser and the following filters: FL1: 530/30, FL2: 585/42, FL3: 670LP, FL4: 660/20, SSC (side scatter): 488/10. The acquired data was analyzed using the CellquestTM software, BD Biosciences.

4.4 Results and discussion

4.4.1 Establishing carotenoid biosynthesis in *S. cerevisiae*

S. cerevisiae strain CEN.PK102-3A was transformed with pSC203 and either pRS405, pSCA101, or pSCA103 to create SimC013, SimC014 and SimC015, respectively; these three carotenoid pigmented strains with characteristic phenotypes can be seen in Figure 4.4.

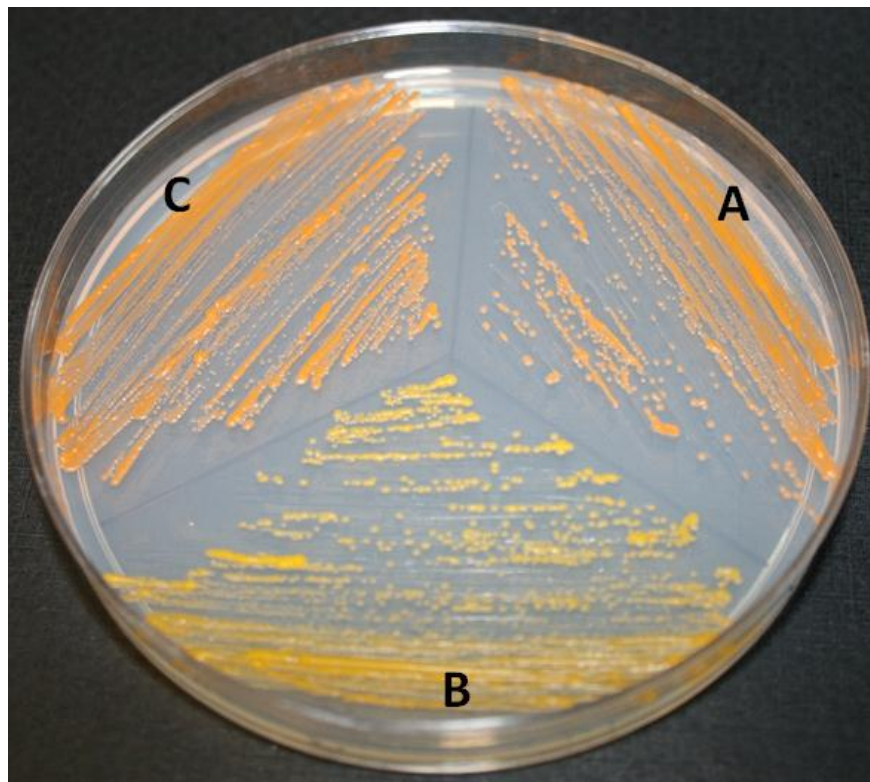


Figure 4.4: Phenotypes of carotenogenic *S. cerevisiae*. A: SimC013 (lycopene producer), B: SimC014 (β -carotene producer) and C: SimC015 (astaxanthin producer).

The lycopene mediated orange-red phenotype of SimC013 closely resembles the astaxanthin phenotype of SimC015, whereas the β -carotene phenotype of the SimC014 stands out with its orange-yellow color. To confirm that the principal carotenoids causing the phenotypes of SimC013, SimC014, and SimC015 are indeed lycopene, β -carotene, and astaxanthin, respectively, the three strains were grown in SC-LEU-URA medium, and the produced carotenoids were extracted into acetone and analyzed by HPLC. The data was compared to HPLC data for authentic carotenoid standards of the three compounds. The results of the HPLC analysis can be seen in Figure 4.5.

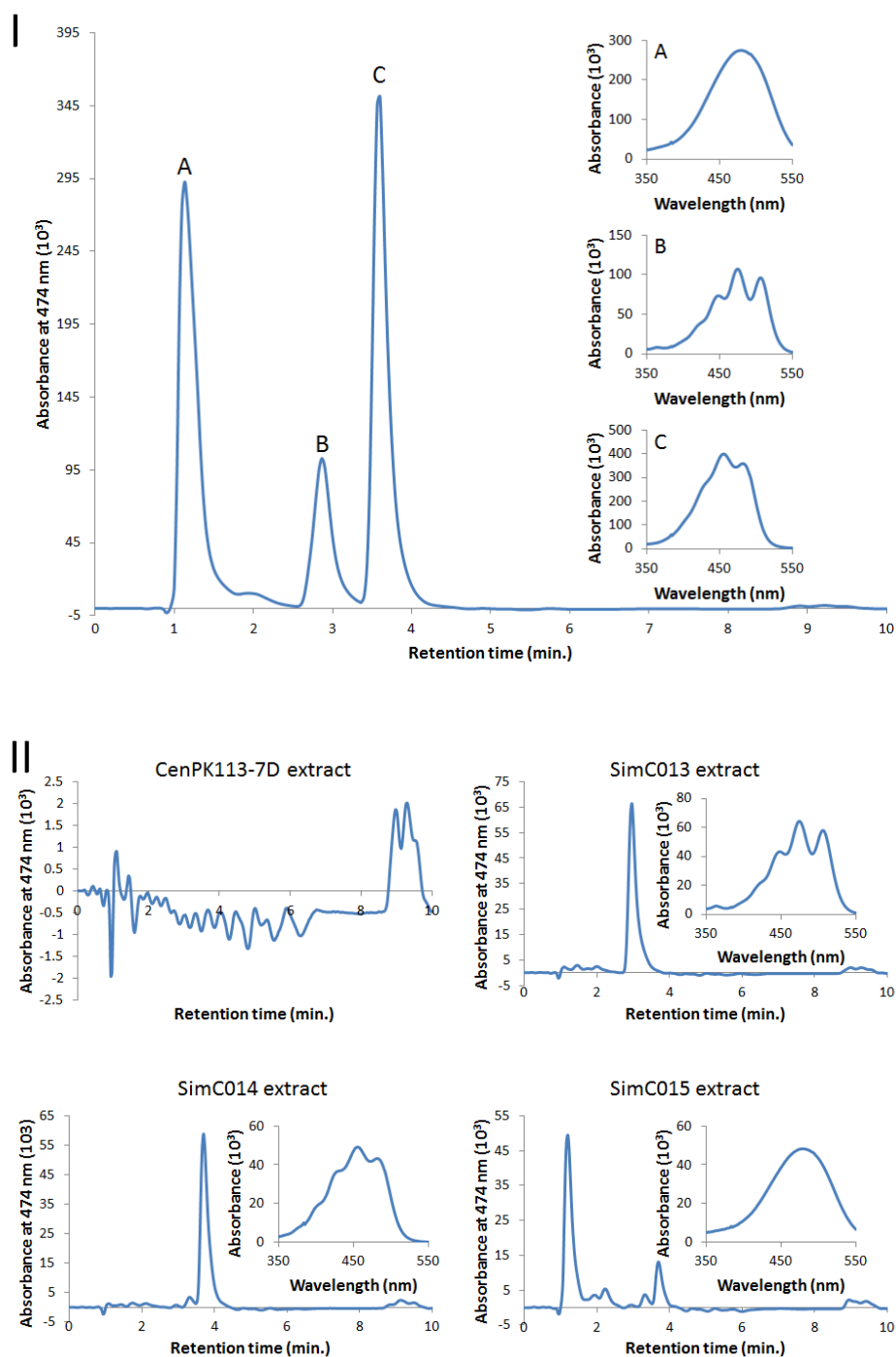


Figure 4.5: HPLC analysis of carotenoids. I: analysis of carotenoid standards in order to determine their retention time and absorbance spectra from 350-550 nm. The chromatogram was recorded at 474 nm. A: astaxanthin, B: lycopene, and C: β-carotene. II: HPLC analysis of acetone extracts from CEN.PK113-7D, SimC013, SimC014 and SimC015. Chromatograms were recorded at 474 nm shown together with the absorption spectrum from 350-550 nm of each of the dominant carotenoid peaks.

Figure 4.5 shows that the retention time of the principal carotenoids produced by SimC013, SimC014, and SimC015 match the retention times of lycopene, β -carotene, and astaxanthin standards, respectively. Furthermore, the UV-spectrum of each carotenoid matches the UV-spectrum of the carotenoid standard, hereby verifying that the produced compounds are identical to the carotenoid standards. The absorption spectra were also verified by comparison to their respective spectra found in the Carotenoids Handbook [Britton et al., 2004]. The HPLC analysis of the SimC014 strain extract shows that there is residual lycopene which means that not all lycopene has been converted to β -carotene, and for the SimC015 strain extract both lycopene and β -carotene are also detected. This shows that the conversion of lycopene to β -carotene and further on to astaxanthin is incomplete. As expected, the extract from the negative control strain CEN.PK113-7D does not show any noticeable carotenoid production.

The carotenoid concentrations in the extracts were quantified and can be seen in Table 4.2.

Table 4.2: Quantification of carotenoids extracted from SimC013, SimC014 and SimC015. The number following the (\pm)-sign is the standard deviations based on biological triplicates.

	Specific titer in mg/l/OD ₆₀₀			
	Lycopene	β -carotene	Astaxanthin	Total
SimC013	0.284 \pm 0.026	ND	ND	0.284
SimC014	0.061 \pm 0.053	0.221 \pm 0.014	ND	0.282
SimC015	0.102 \pm 0.021	0.087 \pm 0.013	0.288 \pm 0.004	0.477

	Yield on biomass in mg/gDCW \dagger (% of total)			
	Lycopene	β -carotene	Astaxanthin	Total
SimC013	1.05 \pm 0.096 (100)	ND	ND	1.05
SimC014	0.225 \pm 0.196 (21.73)	0.815 \pm 0.052 (78.27)	ND	1.04
SimC015	0.376 \pm 0.077 (21.39)	0.321 \pm 0.048 (18.29)	1.06 \pm 0.015 (60.32)	1.76

\dagger Calculated based on the experimentally determined conversion factor for the spectrophotometer (0.271 gDCW/OD₆₀₀/L).

ND = not detectable.

The presence of intermediate metabolites in SimC014 and SimC015 extracts could possibly be eliminated by modulating the expression level of the genes encoding the enzymes catalyzing the individual pathway steps. However, for this study it is sufficient that the carotenoid of interest constitutes the major fraction of the total carotenoid, and the fluorescent phenotype resulting thereof can be considered to result primarily from the principal carotenoid. The buildup of intermediates might create problems later on if it can be shown that one of the carotenoids can be used as a surrogate molecule to identify isoprenoid overproducers, since this might affect the response to changes in upstream metabolites by acting as a buffer system.

The two small peaks appearing adjacent to the large main peak for astaxanthin in the SimC015 extract are probably caused by pathway intermediates between β -carotene and astaxanthin, cf. Figure 4.2. It was not possible to identify the molecules eluting just after astaxanthin since no standards for the intermediate molecules were analyzed using our HPLC method. Even without adding the unidentified carotenoid peaks to the total produced carotenoids SimC015 produces

almost double the amount of carotenoids compared to SimC013 and SimC014. What causes this improvement is unknown, but it could be related to the change in hydrophobicity as a result of the addition of the 3/3'-hydroxyl and 4/4'-keto moieties. However, the total amount of astaxanthin produced by SimC015 is very similar to the total amount of lycopene produced by SimC013 and the total amounts of carotenoids produced by SimC014. This could suggest a capability of *S. cerevisiae* to store higher concentrations of less hydrophobic carotenoids. Verwaal et al. (2010) found that the production of lycopene and β -carotene in *S. cerevisiae* triggered the pleiotropic drug resistance genes, suggesting that carotenoid production confers a stress upon the cells. Their results additionally show that the stress might be membrane stress which corresponds well with the hydrophobic nature of lycopene and β -carotene; this observation, could be related to SimC015's ability to produce a higher total amount of carotenoids.

In a previous study, Verwaal et al. (2007) have shown that *S. cerevisiae* were capable of producing up to 5.9 mg/gDCW β -carotene demonstrating that *S. cerevisiae* is capable of producing and storing higher amounts of β -carotene than what was produced by our strain. It should be noted that this production level was obtained in a strain engineered to overexpress the truncated *HMG1* gene encoding the HMG-CoA reductase and that the carotenoid pathway had been balanced by expression of an extra copy of the *CrtI* gene. Without balancing the carotenoid pathway they were able to produce 0.5 mg/gDCW β -carotene, and their strain without balancing the carotenoid pathway but with the *tHMG1* gene overexpressed produced 1.6 mg/gDCW. These results indicate that we should be able to further increase the β -carotene production if necessary.

Using codon optimized bacterial carotenogenic genes it has been shown that *C. utilis* can be engineered to produce 1.1, 0.4, and 0.4 mg/gDCW lycopene, β -carotene, and astaxanthin, respectively [Miura et al., 1998]. The lycopene amount produced is very similar to the amount we have obtained, whereas β -carotene and astaxanthin is approximately half of what we have obtained. They utilized carotenogenic genes codon optimized for *C. utilis* based on the *crt*-genes from *P. ananatis* and *A. aurantiacum*. This emphasizes the importance of choosing the best genes to base the codon optimized genes upon. We employed genes codon optimized based on the *crtZ* and *crtW* from *Brevundimonas* sp. SD212 which have previously been found to be more effective in heterologous production of astaxanthin in *E. coli*. Ukibe et al. (2009) tested non-codon optimized genes of both yeast (*X. dendrorhous*) and bacterial (*Paracoccus* sp. NBRC101723 and *P. ananatis*) origin. They obtained the best astaxanthin production with the *CrtS* and *CrtR* genes from *X. dendrorhous* (0.12 mg/gDCW). With the bacterial genes they were able to obtain 0.029 mg/gDCW astaxanthin. Since this is the only other example found in literature, our strain is, to date, the best *S. cerevisiae* strain at producing astaxanthin. The best *X. dendrorhous* mutants isolated by FACS by An et al. (1991) produced between 0.39 and 1.25 mg/gDCW total carotenoids (carotenoid distribution has not been specified) whereas the best mutant isolated by FACS by Ukibe et al. (2008) produced 0.50 mg/gDCW astaxanthin. The wild type from which this mutant originates produced 0.13 mg/gDCW astaxanthin. It has hereby been

shown that the SimC013, SimC014, and SimC015 strains produce principally the carotenoids which they were engineered to produce in amounts comparable to the astaxanthin production level of *X. dendrorhous* strains used for FACS. Therefore, these strains served as the basis for testing *in vivo* fluorescence of the three carotenoids in order to determine whether they can be used as surrogate metabolites in FACS screening for isoprenoid overproducing *S. cerevisiae* strains.

4.4.2 Fluorescence microscopy

To test if the constructed strains exhibited fluorescence due to the production of their respective carotenoid, they were subjected to investigation by fluorescence microscopy using the filter normally applied for green fluorescent protein (GFP) and compared to the negative control CEN.PK113-7D. The result can be seen in Figure 4.6.

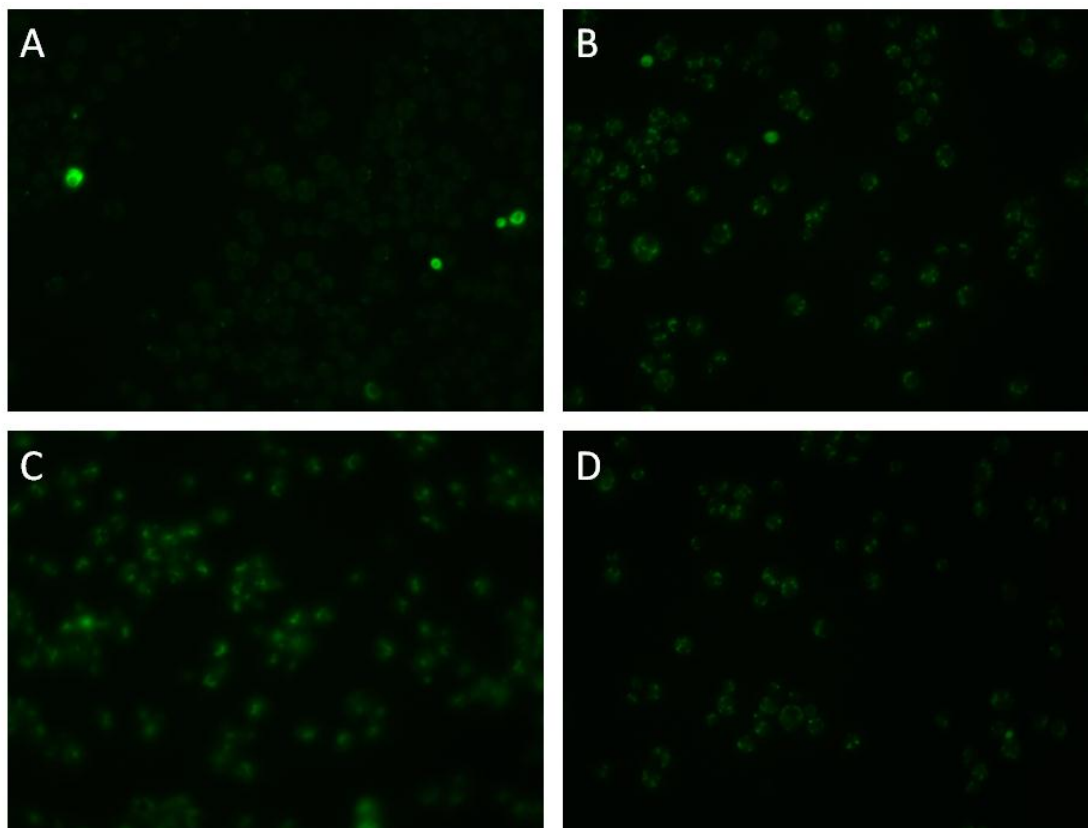


Figure 4.6: Fluorescence microscopy of *S. cerevisiae* cells. A: CEN.PK113-7D, B: SimC013, C: SimC014 and D: SimC015. The pictures were obtained using the Nikon Eclipse TE200-S microscope equipped with 100x objective and GFP band-pass filter (GFP-B: EX480/40, DM505, BA535/50). The exposure time was set to 10 sec.

The filter setup normally applied for GFP was chosen because β -carotene and astaxanthin are excited in the same wavelengths as GFP. This filter setup does however only allow for the capture of fluorescence within the range of the emission filter (BA535/50). This filter value is on the low side of the emission peak for both β -carotene and astaxanthin where the fluorescence is low, cf. Figure 4.1. The data shows that all the carotenogenic strains exhibit fluorescence in small foci within each cell. This finding corresponds well with the finding by Davoli and Weber (2002),

who found that the β -carotene produced by *S. roseus* was stored within lipid droplets. Surprisingly, SimC013 also exhibits fluorescence even though lycopene is not reported to be fluorescent. These results suggest that the fluorescence observed in Figure 4.6: B, C and D are caused by the polyene backbone which all three carotenoids have in common. Carotenoids do in general exhibit very low fluorescence [Davies 1976], which would explain the high exposure time required to obtain the fluorescent images. The high exposure time does of course raise the issue of photo bleaching, but no photo bleaching was observed; at lower exposure times the fluorescence was too weak to obtain proper images. A possibility could be that the fluorescence is caused by other molecules produced by *S. cerevisiae*. However, this is probably not the case since the negative control does not display any fluorescence, and naturally occurring molecules known to exhibit fluorescence such as NAD^+ , NADH, tyrosine and tryptophan do not fluoresce noticeable in the applied wavelengths [Lohmann et al., 1988]. If the fluorescence cannot be attributed directly to the production of the carotenoids it has to be caused by another fluorescent molecule induced by the production of the carotenoids. The most likely explanation is that the fluorescence wavelengths inspected with the filter setup for GFP mainly has to be attributed to the polyene backbone. An et al. (2000) applied confocal laser microscopy to determine the carotenoid concentrations of individual *X. dendrorhous* cells. In this endeavor they applied 488 nm argon ion laser and a 515 LP filter. They found a correlation between total carotenoid content and the carotenoid-mediated fluorescence. Their choice of excitation wavelength corresponds well with our choice of excitation filter, but their emission filter allowed for higher wavelengths to be examined. Our emission filter should be best suited for β -carotene since astaxanthin has a maximum emission at higher wavelengths. We were unable to test higher emission wavelengths at the same excitation wavelengths since the filters for our fluorescence microscope are constructed in fixed sets encompassing an excitation filter and an emission filter. It will be possible to address this issue when analyzing the strains by flow cytometry.

4.4.3 Flow cytometry of carotenoid producing *S. cerevisiae*

The constructed strains together with the negative control CEN.PK113-7D were subjected to analysis by flow cytometry. The analyzed populations were gated based on their forward and side scatter. For each strain, 15,000 occurrences within the gate were recorded. The results can be seen in Figure 4.7.

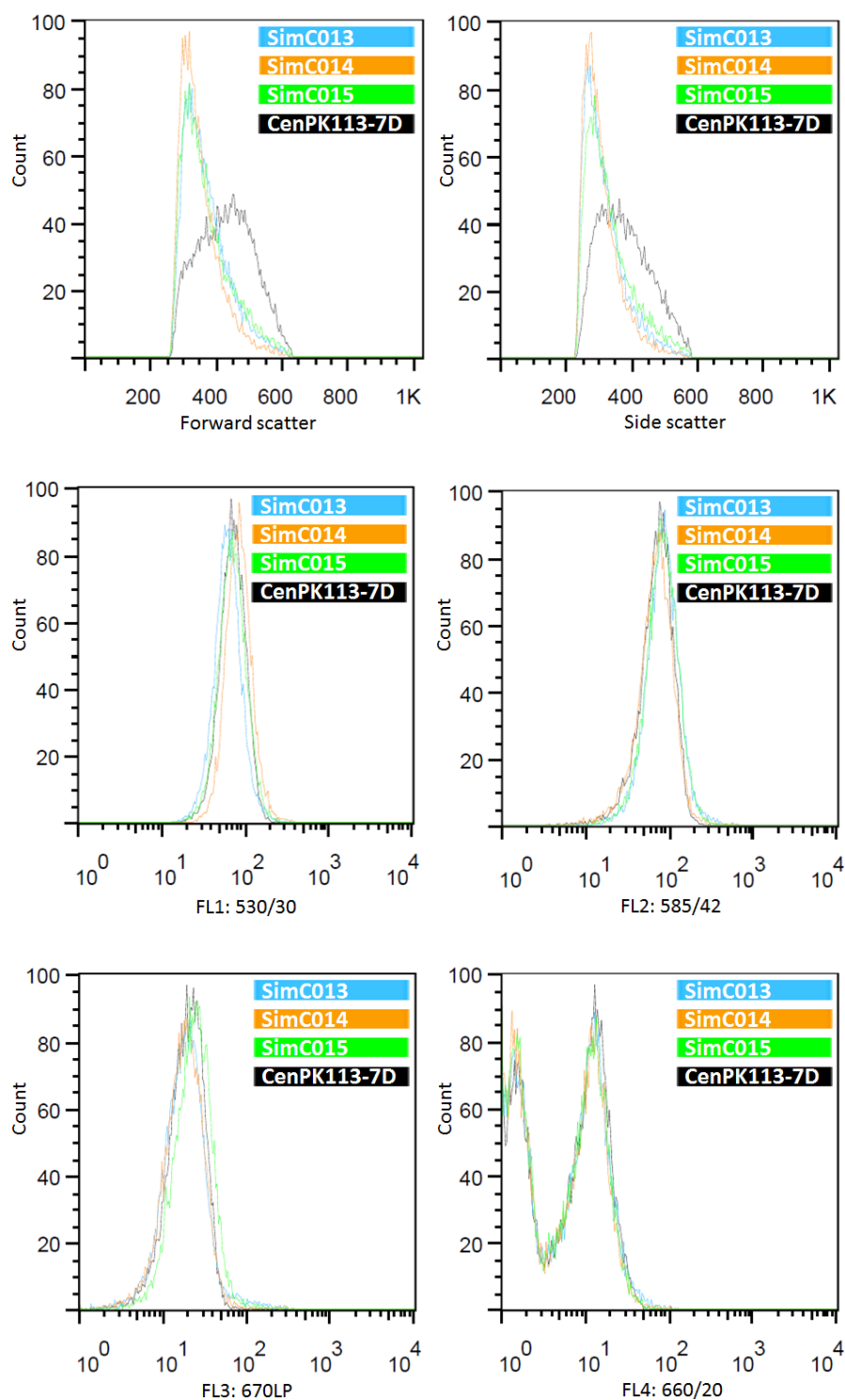


Figure 4.7: Flow cytometry results. SimC013, SimC014, SimC015, and CEN.PK113-7D were subjected to analysis by flow cytometry. The results are shown as histograms for the forward and side scatter together with the fluorescence filtered using the following filters: 530/30, 585/42, 670LP, and 660/20.

From the forward scatter and side scatter histograms, it can be seen that the width of the distributions are the same for all four strains which is due to the populations being gated based on these two parameters. The distribution of the three carotenogenic strains for the forward scatter and side scatter are very similar, whereas the distribution for CEN.PK113-7D is very different from these three. This suggests that the production of carotenoids impact the yeast cell morphology. How carotenoids mediate this effect on cell morphology is unknown.

For the 530/30 filter it can be seen that SimC014 has a distribution with slightly higher fluorescence. This corresponds well with the fluorescence spectra of β -carotene. At this filter SimC015 does not show any difference from CEN.PK113-7D. SimC013 actually has a lower fluorescence than the CEN.PK113-7D which does not correspond with the result obtained from fluorescence microscopy. What the cause is for this decreased fluorescence is unknown, but it could have to do with the light absorbing properties of lycopene. Only very small differences in the distributions of the strains can be observed with the 585/42 filter, which is quite surprising since this is the range of wavelengths where astaxanthin and hereby SimC015 were shown to fluoresce. SimC015 has a bit higher fluorescence than SimC014 and CEN.PK113-7D but a fluorescence distribution similar to SimC013. For the 670LP filter SimC015 stands out from the three other strains with a slight increase in fluorescence. This corresponds well with the shift in fluorescence towards higher wavelengths caused by the keto moieties added to β -carotene in the formation of astaxanthin. No differences between the strains are observed for the 660/20 filter. However, it appears that each strain is encompassed by two discrete populations, namely one with high fluorescence and one with low fluorescence. It is unknown what the cause of this bifurcation is, but it appears to be unrelated to carotenogenesis since it is also exhibited by CEN.PK113-7D. The populations can also be viewed in 2D-cytograms of the aforementioned parameters. We tested all possible combinations of the parameters (data not shown), but none of these allowed efficient discrimination between the strains.

The differences observed between the populations are too minute to distinguish between the strains, hereby making the sorting of cells using FACS almost impossible. The flow cytometer was operated at its most sensitive setting which means that it will be impossible to improve the separation of the population by changing the settings of the voltages and gains of the photomultiplier tubes. FACS has never been applied to carotenogenic *S. cerevisiae*, but examples of its application in selecting carotenoid overproducers from chemical mutagenesis libraries of *X. dendrohous* have been published. An et al. (1991) used a 530/30 filter for which they obtained cytograms as functions of forward scatter. This range of wavelengths corresponds best with the fluorescence of β -carotene but overlaps with the lower end of astaxanthin's fluorescence wavelengths. The application of cytograms as functions of forward scatter allowed them to assess the fluorescence as a function of cell size. This is important when screening for carotenoid overproducers because carotenoids are stored intracellularly and larger cells therefore should be able to accumulate higher amounts of carotenoids; however, it does not necessarily mean that

they exhibit a higher biomass yield of the carotenoids. By applying this technique they were able to isolate *X. dendrohous* carotenoid overproducers, but they did not discriminate between the produced carotenoids. Their results correspond with our data showing a higher fluorescence for the β -carotene-producing SimC014 strain when we applied the 530/30 filter. However, we were unable to discriminate between the populations when we plotted our data in a forward scatter: 530/30 cytogram.

Ukibe et al. (2008) applied several filters (FL1: 490-550, FL2: 570-580, FL3: 600-620, and FL4: 665-685 nm) to analyze chemical mutant libraries of *X. dendrohous*. By applying diphenylamine to inhibit the conversion of β -carotene to astaxanthin, they showed that FL1 correlates well with β -carotene content, whereas the astaxanthin content correlated best with the FL4. They then applied a FL1:FL4 cytogram to select for astaxanthin overproducers. This method does not take cell size into account and thereby not directly assess the astaxanthin biomass yield, but rather the conversion of β -carotene to astaxanthin. Their FL1 and FL4 results do correlate well with our FL1 and FL3 results, cf. Figure 4.7. Their method was not direct sorting where only the best cells are obtained but rather a way of enriching the population after which they had to go through a manual selection step where the best astaxanthin producers were picked from agar plates based on their increased coloration. The low fluorescence of carotenoids seems to hamper the application of FACS for efficient sorting of carotenoid overproducers, which is supported by our results, and until an isoprenoid molecule with higher fluorescence is identified, FACS will probably remain solely a way of enriching a library. The ideal situation would be if the biomass yield of the fluorescent compound could be effectively assessed using a forward scatter:fluorescence cytogram and hereby sorted. If this is not possible enrichment by FACS has to be coupled with other methods which in conjunction would allow the efficient screening for isoprenoid overproducers. Based on our results we decided not to pursue this screening method any further since the gain of applying this method would be too small and the method by itself would not be effective enough. It was therefore decided to pursue other methods for screening for isoprenoid-overproducing *S. cerevisiae* mutants.

4.5 References

An GH, Bielich J, Auerbach R, Johnson EA: **Isolation and characterization of carotenoid hyperproducing mutants of yeast by flow cytometry and cell sorting.** Biotechnology (N Y) 1991, 9:70-73.

An GH, Suh OS, Kwon HC, Kim K, Johnson EA: **Quantification of carotenoids in cells of *Phaffia rhodozyma* by autofluorescence.** Biotechnol. Lett. 2000, 22:1031-1034.

Andrews AG, Phaff HJ, Starr MP: **Carotenoids of *Phaffia rhodozyma*, a red-pigmented fermenting yeast.** Phytochemistry. 1976, 15:1003-1007.

Berry A, Janssens D, Humbelin M, Jore JPM, Hoste B, Cleenwerck L, Vancanneyt M, Bretzel W, Mayer AF, Lopez-Ulibarri R, Shanmugam B, Swings J, Pasamontes L: ***Paracoccus zeaxanthinifaciens* sp. nov., a zeaxanthin-producing bacterium.** Int. J. System Evol. Micr. 2003, 53:231-238.

Boussida S, Vonshak A: **Astaxanthin accumulation in the green algae *Haematococcus pluvialis*.** Plant Cell. Physiol. 1991, 32:1077-1082.

Britton G, Liaaen-Jensen S, Pfander H: **Carotenoids handbook.** Birkhäuser Verlag, Basel, Switzerland 2004.

Choi SK, Nishida Y, Matsuda S, Adachi K, Kasai H, Peng X, Komemushi S, Miki W, Misawa N: **Characterization of β -Carotene ketolases, CrtW, from marine bacteria by complementation analysis in *Escherichia coli*.** Mar. Biotechnol. 2005, 7:515-522.

Choi SK, Matsuda S, Hoshino T, Peng X, Misawa N: **Characterization of bacterial β -carotene 3,3'-hydroxylases, CrtZ, and P450 in astaxanthin biosynthetic pathway and adonirubin production by gene combination in *Escherichia coli*.** Appl. Microbiol. Biotechnol. 2006, 72:1238-1246.

Das A, Yoon SH, Lee SH, Kim JY, Oh DK, Kim SW. **An update on microbial carotenoid production: application of recent metabolic engineering tools.** Appl. Microbiol. Biotechnol. 2007, 77:505-512.

Davies BH: **Carotenoids**, p. 38-165, in Chemistry and Biochemistry of Plant Pigments. Edited by Goodwin TW. Academic Press, London, 1976.

Davoli P, Weber RWS: **Carotenoid pigments from the red mirror yeast, *Sporobolomyces roseus*.** Mycologist. 2002, 16:102-108.

Gietz RD, Woods RA: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.** Method. Enzymol. 2002, 350:87-96.

Ibrahim SF, van den Engh G: **Flow cytometry and cell sorting.** Adv. Biochem. Engin./Biotechnol. 2007, 106:19-39.

Jørgensen K, Stapelfeldt H, Skibsted LH: **Fluorescence of carotenoids. Effect of oxygenation and cis/trans isomerization.** Chem. Phys. Lett. 1992, 190(5):514-519.

Katsuragi T, Tani Y: **Screening for microorganisms with specific characteristics by flow cytometry and single-cell sorting.** J. Biosci. Bioeng. 2000, 89(3):217-222.

Lange N, Steinbüchel A: **β -Carotene production by *Saccharomyces cerevisiae* with regard to plasmid stability and culture media.** Appl. Microbiol. Biotechnol. 2011, 91:1611-1622.

Lohmann W, Lohmann Ch, Ibrahim M: **Fluorescence spectra of NADH/NAD, kynurenine, tryptophan, and tyrosine.** Naturwissenschaften. 1988, 75:141-142.

Mattanovich D, Borth N: **Applications of cell sorting in biotechnology.** Microb. Cell Fact. 2006, 5:12.

Misawa M, Shimada H: **Metabolic engineering for the production of carotenoids in non-carotenogenic bacteria and yeasts.** J. Biotechnol. 1998, 59:169-181.

Miura Y, Kondo K, Saito T, Shimada H, Fraser PD, Misawa N: **Production of the carotenoids lycopene, β -Carotene, and astaxanthin in the food yeast *Candida utilis*.** Appl. Environ. Microb. 1998; 64(4):1226-1229.

Nishida Y, Adachi K, Kasai H, Shizuri Y, Shindo K, Sawabe A, Komemushi S, Miki W, Misawa N: **Elucidation of a carotenoid biosynthesis gene cluster encoding a novel enzyme, 2,2'- β -hydroxylase, from *Brevundimonas* sp. Strain SD212 and combinatorial biosynthesis of new or rare Xanthophylls.** Appl. Environ. Microb. 2005, 71(8):4286-4296.

Nonomura AM, Coder DM: **Improved phycocatalysis of carotene production by flow cytometry and cell sorting.** Biocatalysis. 1988, 1:333-338.

Rentel C, Strohschein S, Albert K, Bayer E: **Silver-plated vitamins: a method of detecting tocopherols and carotenoids in LC/ESI-MS coupling.** Anal. Chem. 1998, 70:4394-4400.

Sikorski RS, Hieter PA: **A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*.** Genetics. 1989, 122:19-27.

Tyo KE, Zhou H, Stephanopoulos GN: **High-throughput screen for poly-3-hydroxybutyrate in *Escherichia coli* and *Synechocystis* sp. Strain PCC6803.** Appl. Environ. Microb. 2006, 72(5):3412-3417.

Ukibe K, Katsuragi T, Tani Y, Takagi H: **Efficient screening for astaxanthin-overproducing mutants of the yeast *Xanthophyllomyces dendrorhous* by flow cytometry.** FEMS Microbiol. Lett. 2008, 286:241:248.

Ukibe K, Hashida K, Yoshida N, Takagi H: **Metabolic engineering of *Saccharomyces cerevisiae* for astaxanthin production and oxidative stress tolerance.** Appl. Environ. Microb. 2009, 75(22):7205-7211.

Vershinin A: **Biological functions of carotenoids – diversity and evolution.** BioFactors 1999, 10:99-104.

Verwaal R, Wang J, Meijnen JP, Visser H, Sandmann G, van den Berg JA, van Ooyen AJJ: **High-level production of beta-carotene in *Saccharomyces cerevisiae* with carotenogenic genes from *Xanthophyllomyces dendrorhous*.** Appl. Environ. Microb. 2007, 73(13): 4342-4350.

Verwall R, Jiang Y, Wang J, Daran JM, Sandmann G, van den Berg JA, van Ooyen AJJ: **Heterologous carotenoid production in *Saccharomyces cerevisiae* induces the pleiotropic drug resistance stress response.** Yeast. 2010, 27(12):983-998.

Visser H, Sandmann G, Verdoes JC: **Xanthophylls in fungi. Metabolic engineering of the astaxanthin biosynthetic pathway in *Xanthophyllomyces dendrorhous*.** In Microbial Processes and Products Methods in Biotechnology™, 2005, 18:257-272. Edited by José-Luis Barredo. Humana Press Inc., New Jersey.

Yamano S, Ishii T, Nakagawa M, Ikenaga H, Misawa N: **Metabolic engineering for production of β -carotene and lycopene in *Saccharomyces cerevisiae*.** Biosci. Biotechnol. Biochem. 1994, 58:1112-1114.

Chapter 5 – Image analysis and lycopene production as tools in semi-high-throughput screening for isopentenyl pyrophosphate overproducing colonies of *Saccharomyces cerevisiae*

Simon Carlsen^{1,2}, Michael Lyng Nielsen^{2,3}, Morten C. Kielland-Brandt², Gregory N. Stephanopoulos¹, Anna Eliasson Lantz^{2,§}

[§]Corresponding author

¹ Department of Chemical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA.

² Department of Systems Biology, Technical University of Denmark (DTU), 2800 Kongens Lyngby, Denmark.

³ Present address: Novozymes A/S, Krogshoejvej 36, DK-2880 Bagsvaerd, Denmark.

Manuscript intended for submission to Microbial Cell Factories

5.1 Abstract

5.1.1 Background

Despite some decades of various chemical and biological approaches to the search for novel pharmaceutical lead compounds, secondary metabolites remain a strong focus for this activity. Most secondary metabolites are only scarcely available in their native production organism, a fact that can be resolved by production in engineered microbes. Fortunately, most secondary metabolites are biosynthesized from very few precursor metabolites. The largest group of secondary metabolites known today is the isoprenoids, encompassing more than 55,000 identified compounds. The main target in optimizing microorganism for production of isoprenoids is to generate microorganisms with a high flux towards the ubiquitous isoprenoid precursor isopentenyl pyrophosphate (IPP). In this endeavor the isoprenoid surrogate red pigment lycopene has been extensively applied in *Escherichia coli* due to the relative ease of which the red phenotypes caused by its production can be screened. However, visual inspection in this approach is labor intensive and subject to errors caused by individual perception of colony color. In addition, colonies of many microorganisms, e.g., *Saccharomyces cerevisiae*, are less transparent than those of *E. coli*, a fact that further hampers visual inspection.

5.1.2 Results

We have developed a simple image analysis method based on a digital camera for image acquisition and the open source software ImageJ 1.44 capable of identifying *S. cerevisiae* colonies with increased red colony coloration due to increased production of lycopene. We found that the lycopene production correlates with the extracted color intensities using a strain engineered to produce lycopene at various levels as a function of the copper concentration in the medium. We finally used the method to screen 15,000 transformants diversified with a *Saccharomyces kluyveri* cDNA library.

5.1.3 Conclusions

The method presented here is capable of selecting colonies with increased lycopene production from images of agar plates with 300 individual colonies. We expect that the demonstration of effectiveness of this method will lead to the development of novel isoprenoid production host systems and aid in the proof-of-concept for new methods and techniques. We see the method as a first step towards a fully automated screening process for microbes overproducing IPP.

5.2 Background

In view of the limited success of approaches like combinatorial chemistry in providing novel pharmaceutical lead compounds, natural products have received renewed attention as a hunting ground for novel chemical scaffolds that can be modified into the next blockbuster pharmaceuticals. It is realized that there is an increased probability of identifying biologically active compounds from a pool of molecules that during the course of evolution have been selected for based on their ability to interact with biological systems [1,2,3]. The redirected search strategy in combination with the inherent complexity of secondary metabolites, as well as the low amounts available from the native production organisms, has created a number of challenges related to the need of obtaining sufficient amounts of the desired compounds. One of the most promising strategies for their large-scale production is biosynthesis through fermentation with microorganisms engineered with heterologous genes. This approach offers several advantages over chemical synthesis, e.g. environmental friendliness and stereo selectivity [4].

Since evolution has structured the production of secondary metabolites to occur from common ubiquitous precursor metabolites, microorganisms already have most of the metabolic machinery required for heterologous biosynthesis. However, the engineered microorganisms still require optimization of the supply of precursor metabolites available for product biosynthesis, but since the secondary metabolite generally does not give any growth advantage, selection of improved production phenotypes becomes difficult or infeasible. In fact, heterologous production of secondary metabolites frequently results in a growth disadvantage, since it removes precursors needed elsewhere in the cell, thereby creating a metabolic burden. Even though selection is preferred over screening, the absence of secondary metabolites with selectable properties necessitates the use of surrogate compounds that can facilitate screening [5]. A microorganism can through genetic modifications and screening based on the properties of the surrogate compound be converted into a production platform organism with a high metabolic flux towards a common precursor that can subsequently be diverted to the biosynthesis of the actual product of interest. A wealth of methods for creating genetic modifications and diversity has been developed, which encompasses both combinatorial and random approaches [6]. However, within this field the use of screens based on surrogate compounds is less developed.

One of the most often applied isoprenoid surrogate compound is lycopene, which owes its red color to the 11 conjugated double bonds absorbing light within the visual range yielding the red phenotype of microorganisms producing it [7]. Lycopene has as a reporter metabolite been instrumental for the success of several metabolic engineering studies, proof-of-concept studies and development of new methods. Lycopene has been applied in studies spanning all the way from random mutagenesis of single protein targets [8], over the effect of chromosomal mutations on plasmid copy number [9], stoichiometric modeling [10], combinatorial effects of deletions and overexpression targets [11], to the screening of gDNA- [12], cDNA- [13], transposon

mutagenesis- [14], and global transcription factor engineering-libraries [15]. All the aforementioned studies were performed in *E. coli* by visual inspection of the degree of red colony coloration caused by changes in the amount of lycopene produced. Lycopene and other carotenoids have also been heterologously produced in *S. cerevisiae* [16], but they have to our knowledge not been used for identification of novel targets for improving isoprenoid production in this organism. The reason may be that *S. cerevisiae* forms more opaque colonies, in which it is more difficult to visually assess the lycopene coloration level. Therefore a more sensitive method for assessing colony coloration becomes imperative.

We here present a simple image analysis method for identifying *S. cerevisiae* colonies with increased red pigmentation due to increased production of lycopene. The method is based on acquisition of digital images of yeast colonies on agar plates with a standard digital camera and analysis with the open source image analysis software ImageJ 1.44 [17]. Besides showing the relationship between the yield of lycopene on biomass yield and the colony coloration calculated by the software, we also apply this method for screening of a *Saccharomyces kluyveri* cDNA library expressed in *S. cerevisiae*.

5.3 Results

5.3.1 Construction of a yeast strain capable of controlled lycopene production

To be able to vary the levels of lycopene produced by *S. cerevisiae*, we constructed strain SimC001, in which the three genes encoding the lycopene pathway are integrated in the *URA3* locus with the first step, geranylgeranyl pyrophosphate (GGPP) synthase, encoded by the *CrtE* gene, under control of the Cu^{2+} inducible *CUPI* promoter. The strain was tested for its ability to produce increasing amounts of lycopene as a function of the Cu^{2+} concentration by streaking it out on SC-URA plates containing different concentrations of copper sulfate, between 0.25 and 200 μM . Colonies were compared to those of the prototrophic non-lycopene producing isogenic control strain CEN.PK113-7D. Figure 5.1 shows that the red colony coloration caused by lycopene indeed increases as the copper sulfate concentration is increased. Furthermore, it can be seen that the colony color of the control strain changes from white to brownish at the highest copper sulfate concentration. This may potentially interfere with the color intensity assessment, since for the SimC001 the coloration at this concentration would be a result of both the lycopene and the stress induced coloration. However, the SimC001 strain should enable us to establish a relationship between the lycopene yield on biomass and the colony coloration.

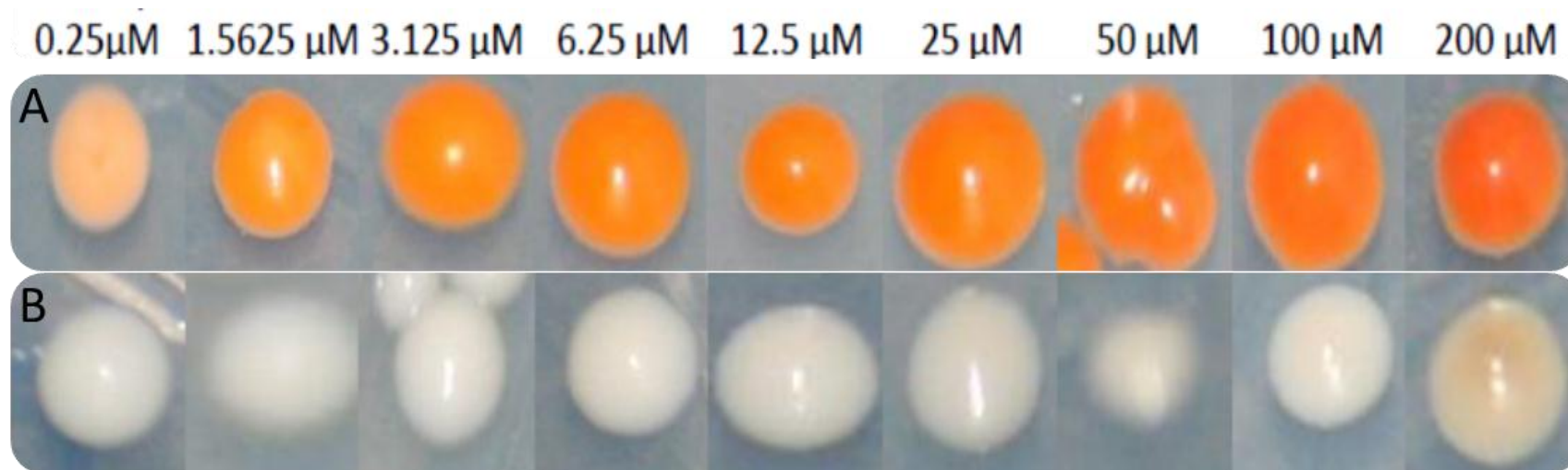


Figure 5.1 - Induced changes in lycopene mediated colony coloration. Colonies of SimC001 (A) and CEN.PK113.7D (B) grown on SC-URA with different copper sulfate concentrations. The copper sulfate concentrations in the agar plates applied are indicated above the respective colony-pictures.

5.3.2 Image acquisition and analysis

To compare the color intensities in digital images of yeast colonies growing on agar plates we made a simple setup based on a standard digital camera, camera mount, black cardboard background and extra ambient lighting (Figure 5.2). This setup allowed us to acquire images of the Petri plates under constant conditions.

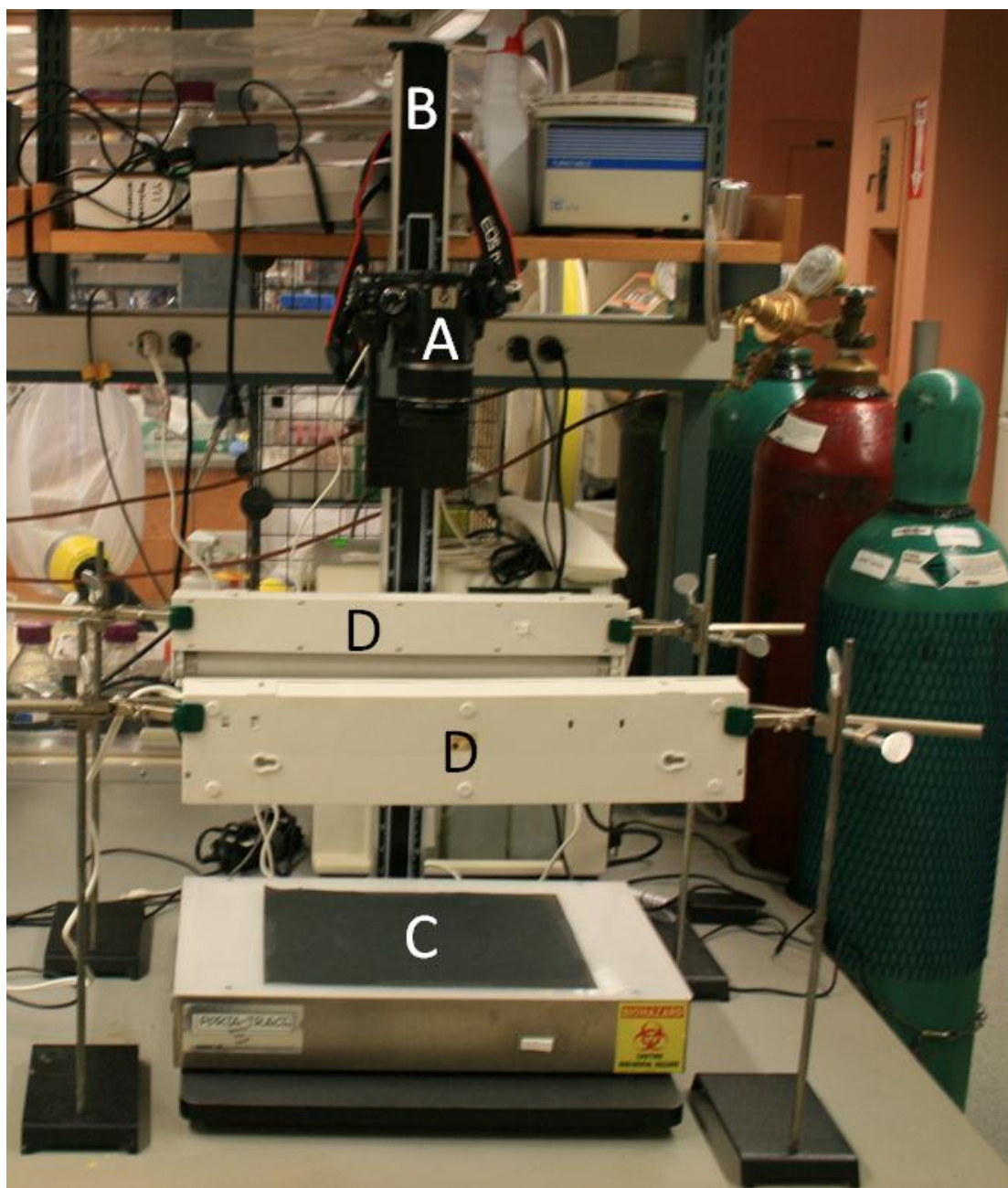


Figure 5.2 - Camera setup. Setup for acquisition of images of agar plates. A, digital camera; B, camera mount; C, black background; D, lights.

To generate agar plates with approximately 300 yeast colonies on each, strain SimC001 was grown in liquid SC-URA until $OD_{600} = 1$, diluted and spread on large SC-URA Petri plates (150 mm diameter) with the same copper sulfate concentrations as indicated in Figure 5.1. Plates were incubated at 30 °C for 3 days and photographed. Figure 5.3 shows an example of a plate image and its processing. The images were analyzed using the ImageJ 1.44 software as follows. The original image (Figure 5.3A) was split into its RGB-channels, which resulted in grayscale images (Figure 5.3B, 5.3C and 5.3D), corresponding to the red, green and, blue channels, respectively. Hereafter the green and the blue image were subtracted from the original image, which resulted in an image that in this example can be seen in Figure 5.3E. In order to enable ImageJ to analyze the images a threshold has to be specified, which resulted in final images corresponding to the example in Figure 5.3F.

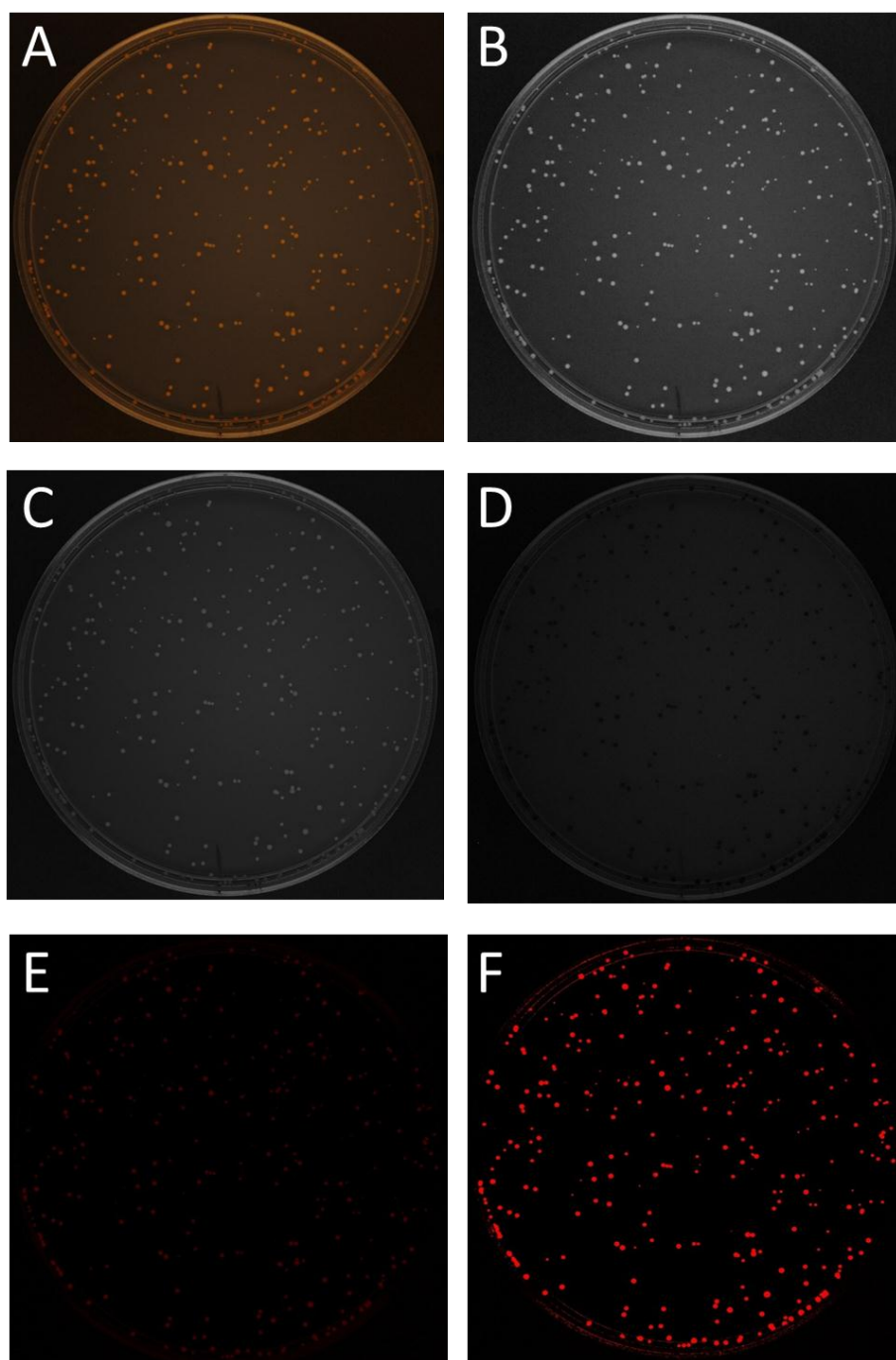


Figure 5.3 - Image processing by ImageJ 1.44. Images showing the steps of the algorithm for processing the original picture to obtain the final picture, which can be analyzed for the red color intensities of the colonies. A, Original image of an agar plate with SimC001 growing on SC-URA + 100 μ M CuSO₄. By splitting the original image into its RGB channels the images corresponding to the red (B), green (C), and blue (D) channels were obtained. Image E was obtained by subtracting the green (C) and blue channel (D) from the original image (A). Image F was obtained by defining the threshold required for further analysis of the image.

Using ImageJ's built-in function called 'Analyze Particles...' we obtained values for the size (in pixels), and the average color intensity (intensity units per pixel) of each colony together with coordinates describing its location (Figure 5.4).

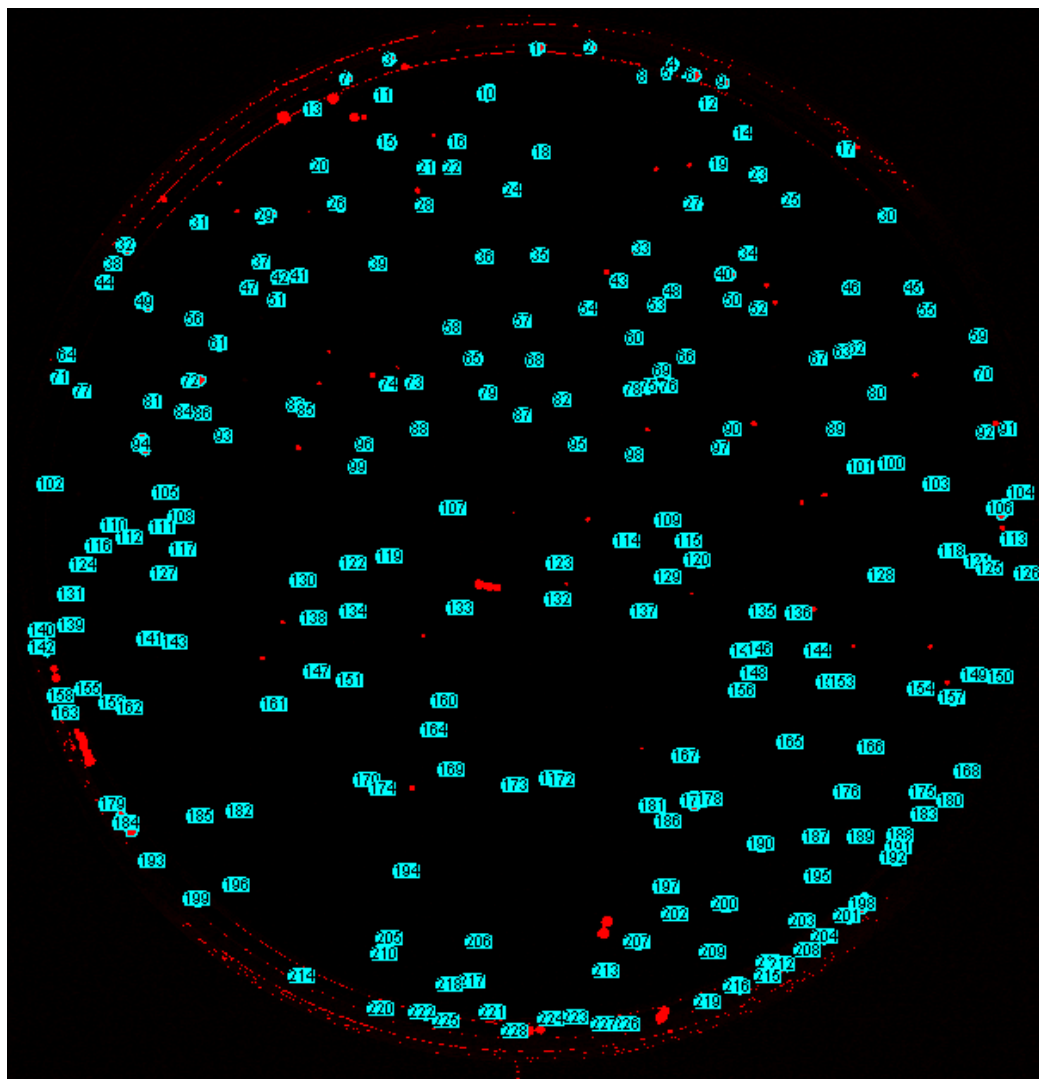


Figure 5.4 - Output image of colonies analyzed for size and red color intensity. After processing to obtain red color intensities, the 'Analyze Particles...' function in ImageJ 1.44 yields an image as the one shown here, where all the analyzed elements are numbered, as well as a table that shows values for size and average red color intensity corresponding to each of the numbered elements.

These operations were performed for all plates with the different copper sulfate concentrations and we calculated the mean of the average color intensity values and standard deviation thereof at each copper sulfate concentration. The results (Figure 5.5A) fit a linear relation between the logarithm of the copper sulfate concentration and the mean of the average red colony color intensities at all concentrations except 200 μ M.

5.3.3 Correlation of the yield of lycopene with red colony color intensity

To correlate colony color intensities with the amount of lycopene produced we grew SimC001 in triplicate liquid SC-URA cultures with each of the copper sulfate concentrations used for the colony color intensity analyses. The strain was grown until stationary phase, and the lycopene was extracted with acetone and quantified by absorbance at 474 nm and converted to concentrations by the use of a standard curve made from pure lycopene dissolved in acetone. Figure 5.5B shows that also the lycopene yield correlates almost linearly with the logarithm of the copper sulfate concentration, except at 200 μ M.

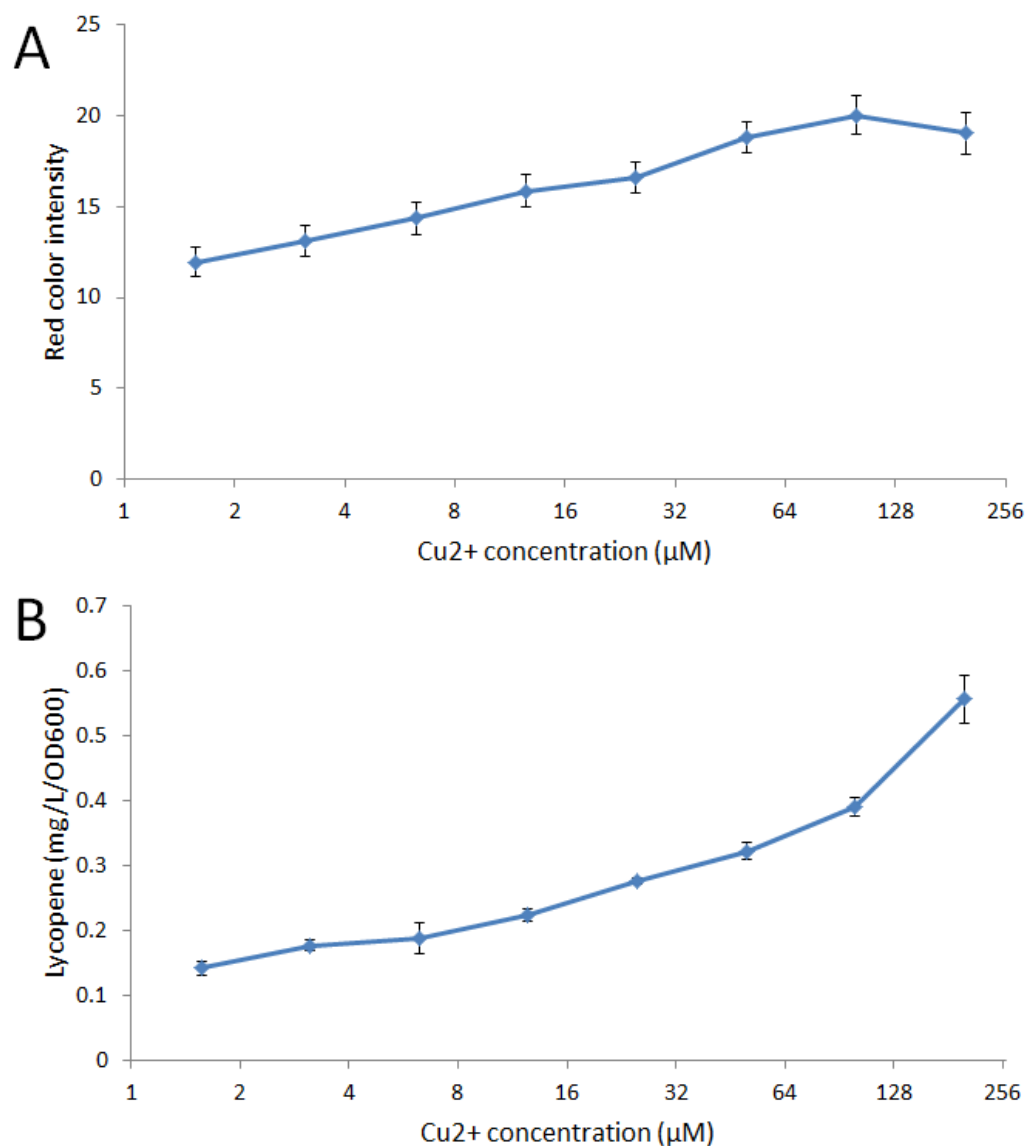


Figure 5.5 - Colony color intensity and lycopene yield as functions of Cu²⁺ concentrations. *S. cerevisiae* strain SimC001 was studied for its ability to produce lycopene and for the associated colony color intensities as functions of Cu²⁺ concentrations in liquid and solid media, respectively. **A:** plates containing copper sulfate at concentrations of 1.5625, 3.125, 6.250, 12.50, 25.00, 50.00, 100.0 and 200.0 μM, respectively, were inoculated with approximately 300 colony forming units of SimC001. Images were recorded and analyzed, and the average red color intensity of all the colonies on each plate was plotted as function of the Cu²⁺ concentration (error bars indicate standard deviation of the mean; the number of colonies analyzed were 287, 326, 376, 358, 358, 358, 337 and 201, respectively). **B:** strain SimC001 was grown in triplicate in 5 ml liquid SC-URA with each of the Cu²⁺ concentrations applied for the agar plates. After all growth had ceased the lycopene concentration normalized with OD₆₀₀ was measured for each culture and plotted as a function of the Cu²⁺ concentration (error bars indicate standard deviation of the mean, n=3).

We plotted the red color intensities of the colonies as a function of the lycopene concentration (Figure 5.6A) and found that a second-order polynomial can approximately describe their relationship. However, one can choose a large sub-interval of lycopene concentrations where this relation can be taken to be linear (Figure 5.6B). These results show that our image analysis method can pick up differences in colony coloration as function of the different amounts of lycopene produced and thus that the method has the potential of being used for screening libraries for colonies with increased lycopene production.

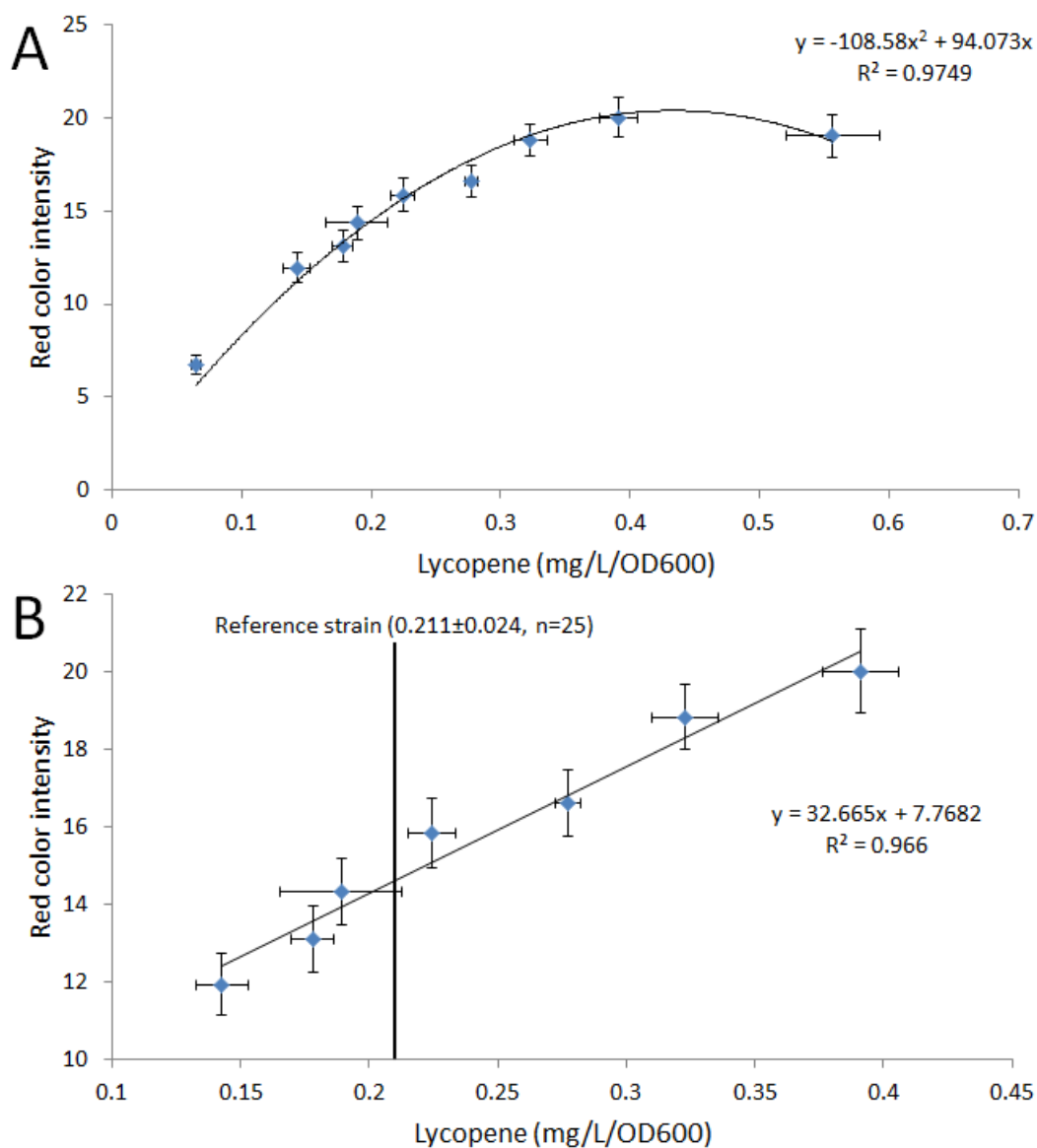


Figure 5.6 – Correlation of colony color intensity with lycopene yield. Red colony color intensity and lycopene concentration normalized with OD₆₀₀ were plotted against each other for each of the eight Cu²⁺ concentrations. Error bars indicate standard deviations of the means. **A:** entire interval of tested values fitted to a second order polynomial relation between red color intensities and lycopene concentrations normalized with OD₆₀₀. **B:** After removal of the outermost value from each side the relation can be fitted to a linear equation. The vertical line indicates the value of the reference strain SimC002 transformed with the pRS414TDH3 control plasmid.

5.3.4 cDNA library screening

In order to challenge our method in an actual library screening, we constructed a cDNA library from *S. kluyveri*, inserted in a single-copy vector under control of the *TDH3* promoter. The plasmid furthermore carried the *TRP1* gene for auxotrophic selection. To test the quality of the library, 20 *E. coli* colonies were picked during the library construction, whereof 14 were found, by sequencing, to harbor a plasmid containing *S. kluyveri* cDNA (result not shown). Furthermore, the library was tested in *S. cerevisiae* by transforming strain CEN.PK113-5D with the cDNA library; by plating the transformants on SC-URA we selected 6 colonies from all of which plasmids harboring the *S. kluyveri URA3* gene were identified by sequencing (results not shown).

To investigate the effects of the cDNA library on the production of lycopene in an engineered *S. cerevisiae*, we constructed strain SimC002, which has the *trp1* auxotrophic marker and carries the genes encoding the lycopene pathway, inserted in the *URA3* locus, with the first step under control of the constitutive *TEF1* promoter. Strain SimC002 was transformed with the cDNA library, and the pRS414TDH3 empty vector as a control, and the transformed cells were spread on 55 large (150 mm diameter) Petri dishes containing SC-TRP-URA, 50 plates for the library and 5 for the control. The plates were incubated for 3 days at 30 °C, whereafter they were photographed for analysis as described above. The 5 colonies scoring highest in the analysis were picked from each plate together with 5 randomly picked colonies. These colonies were transferred to new SC-TRP-URA plates and incubated for 2 days at 30 °C. Each isolate was inoculated into 5 ml of liquid SC-TRP-URA and grown at 30 °C for 2 days. The OD₆₀₀ was measured and 5 ml of fresh SC-TRP-URA were inoculated at OD₆₀₀ = 0.1 from each pre-culture. These cultures were grown until all growth had ceased; the OD₆₀₀ was measured, and the lycopene was extracted and quantified. The mean lycopene concentration normalized with OD₆₀₀ for the randomly selected strains transformed with the empty vector (pRS414TDH3) can be seen plotted as a vertical line in Figure 5.6B. Figure 5.7A presents a box-plot of the measured lycopene concentrations normalized with OD₆₀₀ for both the software-selected and the randomly selected control and library strains. It is seen that the software works better than random sampling of colonies. This is more pronounced for the strains transformed with the library than with the empty vector, as expected due to the higher variation. To test how significant these differences are, we conducted a two-sample t-test from which it for the strains transformed with the control plasmid was found that the software selected isolates have a higher mean value with a p-value of 4.65 %, whereas for the library transformed strains as well were found that the software selected isolates have a higher mean than the randomly selected with a p-value of 3.23 · 10⁻³⁵. This shows that to the software selection allows isolation of transformants with a higher yield of lycopene on biomass. Figure 5.7B shows that the colonies selected from the strain transformed with the library using the software in general gives a higher lycopene yield on biomass.

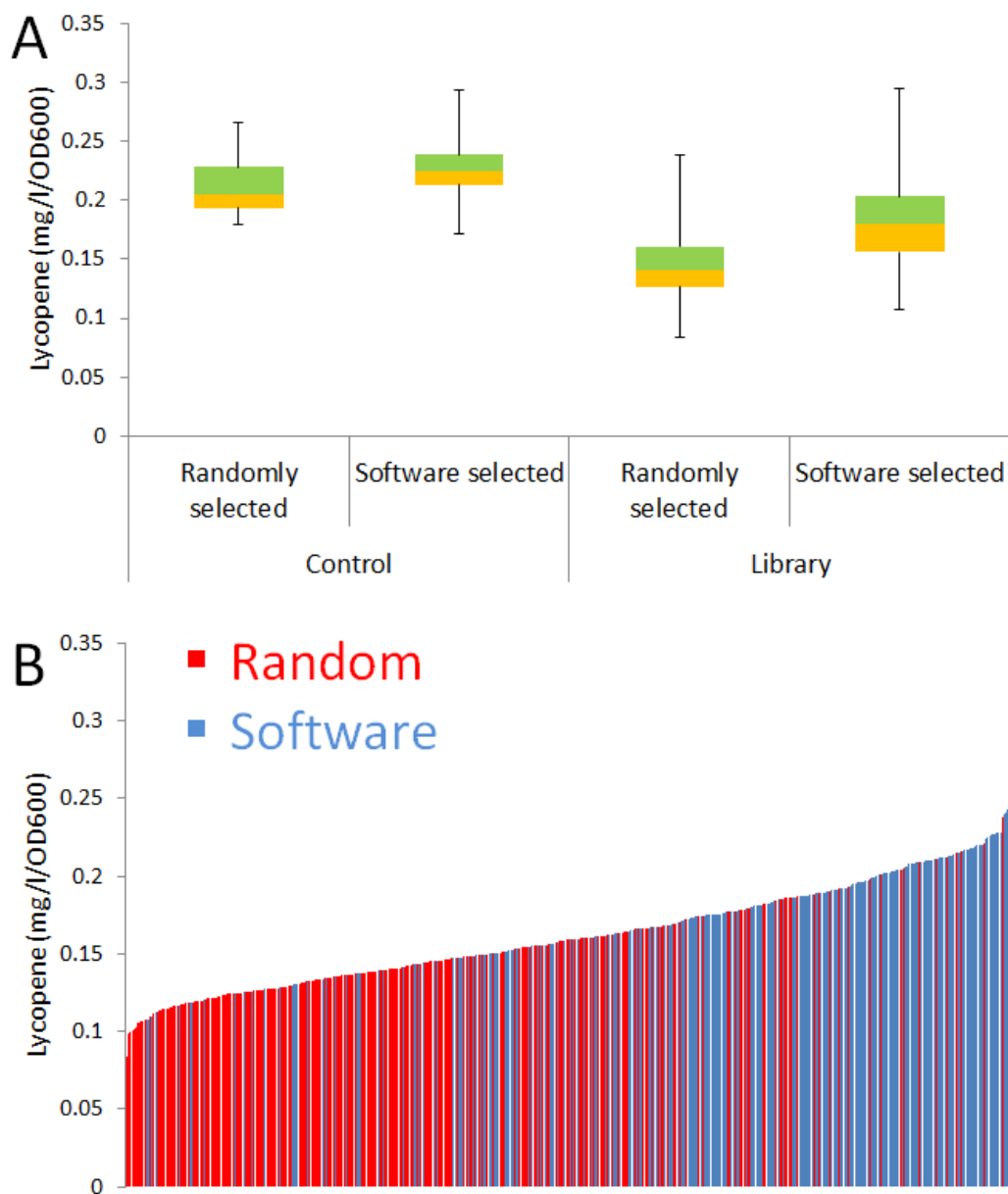


Figure 5.7 – The ability of software-aided selection to pick out *S. cerevisiae* colonies with increased lycopene yield. 25 colonies randomly picked and 25 colonies picked by software-aided selection from the SimC002 strain transformed with the pRS414TDH3 plasmid were analyzed for their level of lycopene production. The same was done for 250 colonies randomly picked and 250 colonies picked by software-aided selection from the SimC002 strains transformed with a *S. kluyveri* cDNA library. **A:** Box-plot of the lycopene concentrations normalized with OD₆₀₀. **B:** Sorted values for the library screen; red bars indicate the values for the colonies randomly selected; blue bars indicate the values for the colonies picked by software-aided selection.

5.4 Discussion

Strain SimC001, capable of producing lycopene at variable levels as a function of the cupric ion concentration in the medium, has helped us investigate the relationship between lycopene production and red colony coloration. However, this strain still contains the *BTS1* gene, encoding *S. cerevisiae*'s native GGPP synthase, which means that the production of lycopene cannot be fully turned off. Nevertheless, this has little impact on the overall strain performance since the activity of the native GGPP synthase is much lower than the codon optimized *CrtE* gene within the range of copper sulfate concentrations relevant in this study. However, it may explain why linear extrapolation of the curves presented in Figure 5.5 does not intersect the y-axis in the origin. It turned out that red color intensities and lycopene concentrations normalized with OD₆₀₀ are linear with the logarithm of copper ion concentration within most of the tested range. A reason for the limited range of linearity could be the characteristic sigmoidal transcription profile of the *CUP1* promoter [18], which would explain the linearity at intermediate Cu²⁺ concentrations and deviations from linearity at both high and low concentrations. This should however not impact the linearity of red colony color intensities as function of lycopene production, since the color intensity should be a function of the produced lycopene. Figure 5.6B shows that an interval of lycopene concentrations normalized with OD₆₀₀ can be chosen for which the red color intensity can be taken to be linear (0.15 – 0.4 mg/l/OD₆₀₀). Verwaal et al. [19] demonstrated that carotenoids produced in high amount in *S. cerevisiae* triggers the pleiotropic drug resistance genes, probably due to membrane stress, which could also impact the relationship between lycopene production and the intensity of red colony coloration. At high lycopene values it would also be expected to find some deviations due to saturation of light absorption.

Figure 5.6B shows that strain SimC002 transformed with a control vector produced lycopene within the linear range and hereby serves as a good starting point for the testing of the ability of our methods to screen a cDNA library. Within this range a response factor can be calculated as the reciprocal value of the slope of the fitted line, i.e., the change in lycopene production required to increase the intensity of red colony coloration by 1 intensity unit, calculated to be 0.0306 mg/l/OD₆₀₀. We tested the ability of the method to screen a population of lycopene producing yeast strains either transformed with the empty vector or the *S. kluyveri* cDNA library made with the same vector and indeed found a higher variation for the library, which should create a more diverse population, than for the control plasmid, which should give a more homogenous population (Figure 5.7A and the result of the t-test). Even though the cells within the control population are all transformed with the same plasmid, small differences do occur. The fact that we with the software were able to pick 25 colonies from this transformation that had a slightly higher mean lycopene production than 25 randomly picked colonies shows the importance of clonal selection. For the library transformants it was expected that most of the plasmids would either have a negative effect or no effect on lycopene production, which was in

agreement with the lycopene production values shown on the box-plot (Figure 5.7A). The t-test shows with a very low p-value that the mean lycopene production for the colonies picked using the software from the library transformants is definitely higher than the ones picked randomly. From Figure 5.7B it is also clear that the majority of the colonies picked by the software are in the high end, whereas the majority of the randomly picked colonies give low production. In general, the highest lycopene producers were identified using the software. However, we were only able to identify mutants with a marginal increase in lycopene production from the cDNA library over the control strain. This was probably due to the lycopene pathway being limiting, as the experiments with the *CUPI* promoter showed that a higher lycopene production can be obtained, and suggested that the *CUPI* promoter at high Cu^{2+} concentrations was stronger than the *TEF1* promoter. Whether this limitation can be alleviated by increasing the copy number by e.g. using a 2 μ -based vector for the expression of the genes encoding the lycopene pathway was not investigated.

Image analysis has been used in a wide array of applications from determination of heterogeneity of cultured strawberry-cells [20] to identification of *Penicillium* species [21]. However, image analysis has not yet gained footing within surrogate compound screening for identification of secondary metabolite overproducing microorganisms, but with high resolution digital cameras becoming a commodity, and with high availability to image analysis software, simple approaches like the method we have presented here will definitely become more widespread.

Visual screens with lycopene as surrogate molecule have been extensively applied; cf. introduction part of this article, but they are laborious, requiring inspection of many agar plates to assess the lycopene mediated colony coloration. With the method presented here screening is faster, and one does not have to worry about the subjective color perception of the person performing the screen. Nevertheless, our method cannot be considered a high-throughput method. We screened approximately 15,000 transformants from the cDNA library, which even with this method is a laborious task. However, the method should be easily automated, and if combined with a colony picking robot, it would result in a true high-throughput screen.

Another screen which relies on colony coloration is for the primary metabolite L-tyrosine. In that screen the L-tyrosine is converted to the black pigment melanin, which enables visual identification of over-producing *E. coli* colonies [22]. Such screens can be adapted with the image analysis method presented here, thereby improving the throughput.

Many colored metabolites can be found in nature and many of these could potentially serve as a surrogate molecule for screening for increased precursor availability of a certain group of secondary metabolites. The method presented here should be applicable to most metabolites that give a change in colony color correlating with its production level when expressed in a colony forming microorganism could potentially be screen via our method.

5.5 Conclusions

We have shown that the method presented here can be applied for screening a population of microbial colonies producing the red pigment lycopene and hereby selecting the individual colonies with increased red coloration. The method could easily be automated and could therefore serve as the initial step towards a fully automated screen. Lycopene has previously demonstrated its worth as a surrogate molecule in several manual screens by visual inspection of *E. coli* colonies, and with the method presented here this approach can easier be applied not only for *E. coli* but in a wide range of microorganisms.

5.6 Methods

5.6.1 Enzymes, chemicals and media

All enzymes were purchased from New England Biolabs Inc., Ipswich, MA and all chemicals from Sigma-Aldrich®, St. Louis, MO, unless otherwise stated. *E. coli* was grown in BD Difco™ LB broth from BD Diagnostic Systems, Sparks, MD, and for plasmid selection ampicillin was added to a final concentration of 100 mg/l. *S. cerevisiae* was grown on SC medium with the appropriate nutrient(s) omitted. This medium was composed of 6.7 g/l Yeast Nitrogen Base (YNB) without amino acids, and also without copper for the copper induction experiment, Sunrise Science Products, Inc., San Diego; CSM – amino acid dropout mixture (amount according to manufacturer's recommendation), Sunrise Science Products, Inc., San Diego, CA; and 20 g/l dextrose. In order to make solid medium for the copper induction experiment, SeaKem LE agarose, Lonza Group Ltd. Basel, Switzerland was applied to avoid copper contamination from regular agar. *S. kluyveri* was grown in YPD, composed of 10 g/l Bacto™ yeast extract, BD Diagnostic Systems; 20 g/l Bacto™ peptone, BD Diagnostic Systems; and 20 g/l dextrose.

5.6.2 Plasmids and strains

All plasmids and strains applied in this study are presented in Table 5.1. The *E. coli* strains listed therein were only used for cloning purposes. Plasmid pSC203 was obtained from the plasmids collection at Center for Microbial Biotechnology at the Technical University of Denmark (DTU). This plasmid is based on the pRS406 plasmid from Mumberg et al., 1995 [23] and contains cassettes for expression of the three genes encoding the lycopene biosynthesis, namely the *TEF1p-CrtE-ADH1t*, *PGK1p-CrtB-ACT1t* and *TDH3p-CrtI-CYC1t*. The three genes were codon-optimized for *S. cerevisiae* and synthesized by Entelechon GmbH, Regensburg, Germany, based on the following protein sequences from *Pantoea agglomerans*, CrtE (GGPP synthase) - GenBank: AAA24819.1, CrtB (phytoene synthase) - GenBank: AAA24821.1 and CrtI (phytoene dehydrogenase) - GenBank: AAA24820.1. Plasmid pRS414TDH3 was obtained as starting point for the construction of a *S. kluyveri* cDNA library.

Strain SimC001 was constructed by transforming CEN.PK113-5D with plasmid pSC203 which was integrated in the *ura3* locus and in the same time had the plasmid based *TEF1* promoter substituted with the *CUP1* promoter. This operation was performed using homologous recombination directly in *S. cerevisiae*, with three DNA fragments. The first two fragments were obtained by digesting pSC203 with StuI and ApaI endonucleases followed by purification by gel-electrophoresis. The StuI cleaves the *URA3* marker to facilitate the genomic integration and the ApaI cleaves just downstream of the *TEF1* promoter. The third DNA fragment was amplified from CEN.PK113-7D gDNA by PCR with primers containing adaptamers homologous to the

backbone of pSC203 just upstream of the *ApaI* recognition site and downstream of the *TEF1* promoter, respectively. It was verified by PCR that this transformation resulted in the integration of the three joined fragments in the *ura3* locus of CEN.PK113-5D and the exchange of the *TEF1* promoter with the copper inducible *CUP1* promoter. Strain SimC002 was constructed by transforming CEN.PK 113-9D with pSC203 which prior to the transformation had been digested with endonuclease *StuI* and purified by gel electrophoresis.

Table 5.1 - Strains and plasmids. Acquired and constructed strains and plasmids applied in this study.

Strain	Genotype	Source
<i>Saccharomyces cerevisiae</i>		
CEN.PK113-7D	<i>MATα MAL2-8^c SUC2</i>	Dr. P. Kötter, Frankfurt, Germany.
CEN.PK113-5D	<i>MATα MAL2-8^c SUC2 ura3</i>	Dr. P. Kötter, Frankfurt, Germany.
CEN.PK113-9D	<i>MATα MAL2-8^c SUC2 trp1 ura3</i>	Dr. P. Kötter, Frankfurt, Germany.
SimC001	<i>MATα MAL2-8^c SUC2 ura3::(<i>CUP1p-CrtE-ADH1t</i> <i>PGK1p-CrtB-ACT1t TDH3p-CrtI-</i> <i>CYC1t URA3)</i></i>	This study.
SimC002	<i>MATα MAL2-8^c SUC2 trp1, ura3::pSC203 (<i>TEF1p-CrtE-</i> <i>ADH1t PGK1p-CrtB-ACT1t</i> <i>TDH3p-CrtI-CYC1t URA3)</i></i>	This study.
<i>Saccharomyces kluyveri</i>		
NRRL12651	Wild type	Dr. Felix Lam and Dr. Gerald Fink, Whitehead Institute, MIT, Cambridge, MA.
<i>Escherichia coli</i>		
MAX Efficiency® DH5 α TM	F- ϕ 80 <i>lacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (rk-, mk+) <i>phoA supE44</i> λ - <i>thi-1</i> <i>gyrA96 relA1</i>	Life Technologies, Grand Island, NY.
ElectroMAX TM DH10B TM T1 Phage-Resistant Competent Cells	F ⁻ <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74 recA1</i> <i>endA1 araD139</i> Δ (<i>ara, leu</i>)7697 <i>galU galK</i> λ - <i>rpsL nupG tonA</i>	Life Technologies, Grand Island, NY.
Plasmid name	Origin/reference	

pSC203	Center for Microbial Biotechnology plasmid collection, DTU*.
pRS414TDH3	Mumberg et al., 1995 [23].

* Constructed by Simon Carlsen as part of his Ph.D. thesis work.

5.6.3 Image acquisition and analysis

Images were acquired applying the setup shown in Figure 5.2. The setup is based on a camera mount to keep the distance between the camera and the agar plate constant, a Canon EOS Rebel XTi digital camera, extra lights to illuminate the agar plate, hereby avoiding the use of a flash, and a black background to minimize light reflected from the background. The camera was used with manual, constant settings. The images were saved in .jpg format.

Images were analyzed with the ImageJ 1.44 software, using the standard functions available, which in the following are referred to with the functions and parameters shown in quotation marks. First the image was split into its red, green and blue (RGB) channels using the “Split Channels” function, and the green and blue images were subtracted from the original image using the “Image Calculator...” function. The resulting image was adjusted using the “Threshold...” function to remove any remaining background light. This was done by changing the threshold of the “Brightness” and keeping the settings for “Hue” and “Saturation” constant. We found that we got the best result with a “Brightness” threshold at 20, but this varies depending on the images so it should be determined for each series of images and hereafter kept constant for the analysis of this series of images. To analyze each of the colonies, we applied the “Analyze Particles...” function. Prior to applying this function we specified the desired output values in the “Set Measurements...” function. Here we checked the following options: Area (area of the analyzed particle), Mean gray value (the mean color intensity of the particle), Standard deviation (standard deviation corresponding to each of the mean color intensities), Shape descriptors (values for circularity, aspect ratio, roundness and solidity) and Limit to threshold (ensuring that only pixels with light intensities above the threshold are analyzed). The area and shape descriptor values are only used to determine the best value for identification of the colonies which needs to be specified in the “Analyze Particles...” function, thus allowing for the analysis of single colonies and the discarding of joined colonies and other artifacts occurring in the image. The main value of interest is the mean gray value, which is the value used directly as the red color intensity of each colony. We found that the following settings gave the best results when we applied the “Analyze particle...” function: Size range of the analyzed particles = 200 – infinity (pixels) and circularity range = 0.7 – 1.0 (circularity is calculated as $4\pi \frac{\text{area}}{\text{perimeter}^2}$, where 0 corresponds to an infinitely elongated polygon and 1 to a perfect circle. We furthermore specified that the function should “show overlay masks” and checked the following boxes:

Display results (opens separate window with a results table), Clear results (removes results from the results table from previous analysis) and Exclude on edges (excludes particle touching the edge of the image from the analysis). After applying the “Analyze Particles...” function the results were transferred from the results table into Microsoft Excel for further analysis. The values were sorted using Excel to identify colonies with the highest red color intensity which was subsequently picked for further analysis. For further information please refer to the ImageJ 1.44 user guide: <http://rsbweb.nih.gov/ij/docs/user-guide.pdf>.

5.6.4 Lycopene extraction and quantification

S. cerevisiae cultures producing lycopene were grown until stationary phase in 5 ml liquid SC medium with the appropriate nutrient(s) omitted. One ml from each culture was transferred to a 2 ml FastPrep® tube, MP Biomedicals LLC, Solon, OH, and spun down at 18,000 g for 2 min. Medium was carefully removed with a pipette, and the cells were re-suspended in 1 ml acetone containing 100 mg/l 2,6-di-*tert*-butyl-4-methylphenol. One fifth ml acid washed glass beads (425-600 µm, Sigma-Aldrich®) was added and the tubes capped. Lycopene was extracted in a Savant FastPrep® FP120 Cell Disrupter, MP Biomedicals LLC, at maximum speed for 3 times 45 sec with 1 min cooling on ice between each run. The acetone extracts were filtered using a 1 ml syringe equipped with a 13 mm, 0.2 µm PTFE filter, VWR, Radnor, PA and quantified by measuring the absorbance at 474 nm and relating the measurement to a standard curve made from authentic lycopene dissolved in acetone with 100 mg/l 2,6-di-*tert*-butyl-4-methylphenol.

5.6.5 cDNA library construction and screening

Total RNA was extracted using the Savant FastPrep® FP120 Cell Disrupter, MP Biomedicals LLC, glass beads (425-600 µm), Sigma-Aldrich®, 500 µl acidic phenol:chloroform:isoamyl alcohol 125:24:1), pH 4.5, Life Technologies, Grand Island, NY, and 500 µl nuclease free water, Qiagen, Valencia, CA, from the cells harvested from 5 ml culture of *S. kluyveri* grown in YPD until OD₆₀₀ = 1. Four hundred and fifty µl of the aqueous phase were transferred to a micro-centrifuge tube. Total RNA was precipitated by addition of 45 µl 3 M sodium acetate, pH 5.2 and 450 µl isopropanol. The RNA was spun down, rinsed twice with 500 µl 80 % ethanol and re-dissolved in 200 µl nuclease free water. mRNA was purified from the total RNA with Oligo d(T)25 Magnetic Beads from New England Biolabs Inc., Ipswich, MA, but instead of eluting the mRNA, the following steps were performed with the mRNA bound to the magnetic beads which aided in the buffer and reagent changes. The protocol was inspired by [24] and [25]. The mRNA was reverse transcribed using the M-MuLV reverse transcriptase and the poly T from the magnetic beads as primer, whereafter T4 DNA polymerase was applied to remove oligo dTs that had not bound mRNA and thereby not been used for primers in the synthesis of cDNA. The

mRNA strands were removed by treating the mRNA/cDNA hybrid mixture with RNase H. The now single stranded cDNA was treated with terminal transferase (TdT terminal transferase, NEB catalog# M0315) and dCTP, hereby creating a 3' poly C tail on the cDNA. The single stranded cDNA bound to the magnetic beads was then applied as template for a standard PCR reaction setup using the Phusion® High-Fidelity DNA Polymerase together with a poly G primer and a poly T primer with adaptamers containing the PacI and AscI endonuclease restriction sites, respectively, which resulted in the formation of double stranded cDNA flanked by the PacI and AscI restriction sites. The cDNA was size separated by gel electrophoresis, recovered in a 1-10 kb size range and treated with the PacI and AscI restriction endonucleases. The cDNA was purified using the PCR purification kit (Qiagen) ligated into the pRS414TDH3 plasmid which had been engineered to harbor the same restriction sites within its multiple cloning site. The plasmid was prior to the ligation treated with the same two restriction endonucleases and purified by gel electrophoresis. The 50 µl ligation mix was precipitated with the addition of 5 µl 3 M sodium acetate, pH 5.2 and 50 µl isopropanol, washed twice with ice cold 80 % ethanol and redissolved in 5 µl nuclease free water, Qiagen. A half microliter of the library was transformed into ElectroMAX™ DH10B™ T1 Phage-Resistant Competent Cells and plated on large (150 mm in diameter) LB + ampicillin plates and incubated overnight at 37 °C. Twenty colonies were picked onto new LB + ampicillin plates for analysis of the library quality. The rest of the colonies were scraped of the plates using 1 ml ice cold LB medium per plate and a Drigalski spatula. Plasmids from the collected cells were purified using the Plasmid Midi Kit, Qiagen and hereafter constituted the library while the remainder of the concentrated ligation mix was stored at -20 °C as a master library in case more library has to be generated.

5.7 List of abbreviations

GGPP - Geranylgeranyl pyrophosphate

IPP - Isopentenyl pyrophosphate

5.8 Competing interests

The authors declare that they have no competing interest.

5.9 Author's contributions

SC conceived the idea for the project, carried out the experiments, analyzed the data and drafted the initial manuscript. MLN, AEL, GNS and MKB provided suggestions, corrections and help with design of the experiments, the analysis of the data, and preparation of the manuscript. All the authors have approved the final manuscript.

5.10 Acknowledgements

The authors acknowledge Dr. Felix Lam and Dr. Gerald Fink for contributing the plasmids and *S. kluyveri* strain applied in this study. Furthermore, the authors are grateful to the Technical University of Denmark for funding. The authors would also like to thank Massachusetts Institute of Technology for funding.

5.11 References

1. Paterson I, Anderson EA: **The renaissance of natural products as drug candidates.** Science 2005, **310**:451-453.
2. Harvey AL: **Natural products as a screening resource.** Curr Opin Chem Biol 2007, **11**:480-484.
3. Demain AL, Fang A: **The natural functions of secondary metabolites.** In Advances in Biochemical Engineering/Biotechnology. Volume 69. History of Modern Biotechnology I, Edited by Fiechter A. Springer-Verlag Berlin Heidelberg; 2000:1-39.
4. Rohlin L, Oh M, Liao JC: **Microbial pathway engineering for industrial processes: evolution, combinatorial biosynthesis and rational design.** Curr Opin Microbiol 2001, **4**:330-335.
5. Zhang YHP, Himmel ME, Mielenz JR: **Outlook for cellulose improvement: screening and selection strategies.** Biotechnol Adv 2006, **24**:452-481.
6. Klein-Marcuschamer D, Ajikumar PK, Stephanopoulos G: **Engineering microbial cell factories for biosynthesis of isoprenoid molecules: beyond lycopene.** Trends Biotechnol 2007, **25**:417-424.
7. Vershinin A: **Biological functions of carotenoids – diversity and evolution.** BioFactors 1999, **10**:99-104.
8. Leonard E, Ajikumar PK, Thayer K, Xiao WH, Mo JD, Tidor B, Stephanopoulos G, Prather KLJ: **Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control.** Proc Natl Acad Sci USA 2010, **107**:13654-13659.
9. Tao L, Jackson RE, Rouvière PE, Cheng Q: **Isolation of chromosomal mutations that affect carotenoid production in *Escherichia coli*: mutations alter copy number of ColE1-type plasmids.** FEMS Microbiol Lett 2005, **243**:227-233.
10. Alper H, Jin YS, Moxley JF, Stephanopoulos G: **Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*.** Metab Eng 2005, **7**:155-164.
11. Jin YS, Stephanopoulos G: **Multidimensional gene target search for improving lycopene biosynthesis in *Escherichia coli*.** Metab Eng 2007, **9**:337-347.
12. Kang MJ, Lee YM, Yoon SH, Kim JH, Ock SW, Jung KH, Shin YC, Keasling JD, Kim SW: **Identification of genes affecting lycopene accumulation in *Escherichia coli* using a shot-gun method.** Biotechnol Bioeng 2005, **91**:636-642.
13. Gallagher CE, Cervantes-Cervantes M, Wurtzel ET: **Surrogate biochemistry: use of *Escherichia coli* to identify plant cDNAs that impact metabolic engineering of carotenoid accumulation.** Appl Microbiol Biotechnol 2003, **60**:713-719.
14. Alper H, Miyaoku K, Stephanopoulos G: **Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets.** Nat Biotechnol 2005, **23**:612-616.
15. Alper H, Stephanopoulos G: **Global transcription machinery engineering: a new approach for improving cellular phenotype.** Metab Eng 2007, **9**:258-267.
16. Yamano S, Ishii T, Nakagawa M, Ikenaga H, Misawa N: **Metabolic engineering for production of β -carotene and lycopene in *Saccharomyces cerevisiae*.** Biosci Biotech Biochem 1994, **58**:1112-1114.

17. Abramoff MD, Magelhaes PJ, Ram SJ: **Image processing with ImageJ.** Biophotonics Int 2004, **11**:36-42.
18. Gorman JA, Clark PE, Lee MC, Debouck C, Rosenberg M: **Regulation of yeast metallothionein gene.** Gene 1986, **48**:13-22.
19. Verwaal R, Jiang Y, Wang J, Daran JM, Sandmann G, van den Berg JA, van Ooyen AJ: **Heterologous carotenoid production in *Saccharomyces cerevisiae* induces the pleiotropic drug resistance stress response.** Yeast 2010, **12**:983-998.
20. Miyanaga K, Seki M, Furusaki S: **Quantitative determination of cultured strawberry-cell heterogeneity by image analysis: effects of medium modification on anthocyanin accumulation.** Biochem Eng J 2000, **5**:201-207.
21. Dörge T, Carstensen JM, Frisvad JC: **Direct identification of pure *Penicillium* species using image analysis.** J Microbiol Meth 2000, **41**:121-133.
22. Santos CNS, Stephanopoulos G: **Melanin-based high-throughput screen for L-tyrosine production in *Escherichia coli*.** Appl Environ Microb 2008, **74**:1190-1197.
23. Mumberg D, Müller R, Funk M: **Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds.** Gene 1995, **156**:119-122.
24. Roeder T: **Solid-phase cDNA library construction, a versatile approach.** Nucleic Acids Res. 1998, **26**:3451-3452.
25. Lambert KN, Williamson VM: **Library construction from small amounts of RNA using paramagnetic beads and PCR.** In: cDNA Library Protocols, Methods in Molecular Biology, volume 69. Edited by Cowell IG, Austin CA. Totowa, New Jersey: Humana Press Inc. 1997:1-12.

Chapter 6 – Teaching old DNA new tricks: Engineering an *in vivo* transposon mutagenesis system for *Saccharomyces cerevisiae* based on the bacterial Tn5-transposase

6.1 Abstract

Transposons have been modified to create novel tools for molecular biology and genetics. Several systems have been developed for transposon mutagenesis of *Saccharomyces cerevisiae*, but all of these suffer from drawbacks limiting their usefulness. We here demonstrate the design and construction of a novel transposase, based on the Tn5 transposase, which should exhibit increased functionality in *S. cerevisiae*, hereby remedying the drawbacks of previous systems. Our results show that this novel transposase locates to the nucleus of *S. cerevisiae*. Furthermore, we applied this new version of the transposase to develop a novel plasmid-based tool for random *in vivo* mutagenesis of *S. cerevisiae*. The construction of this plasposon system and its control plasmids are described here together with the initial testing of the system. The preliminary results for the testing of the system suggest that it is functional. However, further experiments are required to demonstrate that the system works correctly and performs the task for which it was designed.

6.2 Introduction

Since the first transposon was discovered in maize more than 60 years ago by Barbara McClintock [McClintock, 1950], for which she in 1983 was awarded the Nobel prize, transposons have been developed as versatile tools for genetic engineering and molecular biology [Vizváryová and Valkov, 2004]. The two most commonly applied types of transposon tools for mutagenesis of *S. cerevisiae* are the systems based on one of the native Ty-retro-transposons [Curcio and Garfinkel, 1991] and the shuttle mutagenesis systems utilizing bacterial transposons [Seifert et al., 1986]. Of the two, the Ty retro-transposition system is the only one in which the actual transposition event takes place *in vivo* in *S. cerevisiae*; the shuttle mutagenesis system relies on the transposition to take place *in vivo* in *E. coli* into cloned DNA fragments which subsequently can be shuttled back to *S. cerevisiae* where chromosomal insertions are mediated by homologous recombination. As such, the drawback of shuttle mutagenesis is the requirement for construction of a library of cloned gDNA fragments before the transposition mutagenesis can be carried out. The Ty based systems also suffer from several drawbacks, e.g. the transpositions have been found to occur into non-random targets [Natsoulis et al., 1989], especially with integration preference upstream of genes transcribed by RNA polymerase III [Devine and Boeke, 1996]. Furthermore, as native transposon elements in *S. cerevisiae*, the Ty elements, and scars from them being looped out of the genome, can be found throughout all the chromosomes of *S. cerevisiae* [Kim et al., 1998]. This creates a bias towards loci-specific insertion by homologous recombination instead of transposition.

To remedy these problems, several *in vitro* transposition systems based on bacterial transposons have been created (for a comprehensive list see Smith, 2007). These systems can be applied in mutagenesis of yeast by transposition of the desired DNA into randomly fragmented yeast gDNA followed by transformation of the yeast gDNA donor. However, these systems rely on purified transposase enzymes which have to be purchased or manually purified. It would be desirable to have a system independent of *in vitro* transposition and thus the requirement to obtain purified transposase enzyme. A solution for this problem could be the construction of a plasmid-based system both carrying both the gene encoding the transposase and the DNA to be delivered into the genome of the target organism. Such a system has been developed for *E. coli* based on the bacterial Tn5 transposase. Dennis and Zylstra (1998) constructed what they coined as a plasposon system. The system's main components were the gene encoding the Tn5 transposase, the transposase 19 bp recognition sequences, an antibiotic selection marker and a conditional origin of replication. A system like this has been created for yeast but only based on its native Ty transposon system, which as mentioned above suffers from several drawbacks.

In this work, we have pursued the construction of a plasposon system for yeast based on a bacterial transposon. We chose to base our system on a hyper-active version of the Tn5 transposase, since the Tn5 system requires only its transposase and 19 bp recognition sequences for activity, whereas other bacterial transposons require either longer recognition sequences or

multiple components for functionality (discussed by Goryshin and Reznikoff, 1998). Furthermore, it has been demonstrated that upon transformation of *S. cerevisiae*, *in vitro*-generated synaptic complexes, composed of Tn5 transposase dimers and DNA flanked by the 19 bp recognition sequences, could be transposed into the genome [Goryshin et al., 2000].

Tn5 belongs to the class II ‘cut and paste’ transposons, and its transposition is composed of a multistep process. The Tn5 transposase first binds to its 19 bp recognition sequence, followed by dimerization to form the synaptic complex; next, the synaptic complex is cleaved from the donor DNA and inserted into its 9 bp target which is cleaved by a staggered cut and presumably repaired by host DNA repair proteins such as to create a 9 bp sequence duplication on each side of the insert [Reznikoff, 2002]. A schematic representation of this process can be seen in Figure 6.1. For a more comprehensive description of the Tn5 transposon mechanism see Reznikoff, 2002.

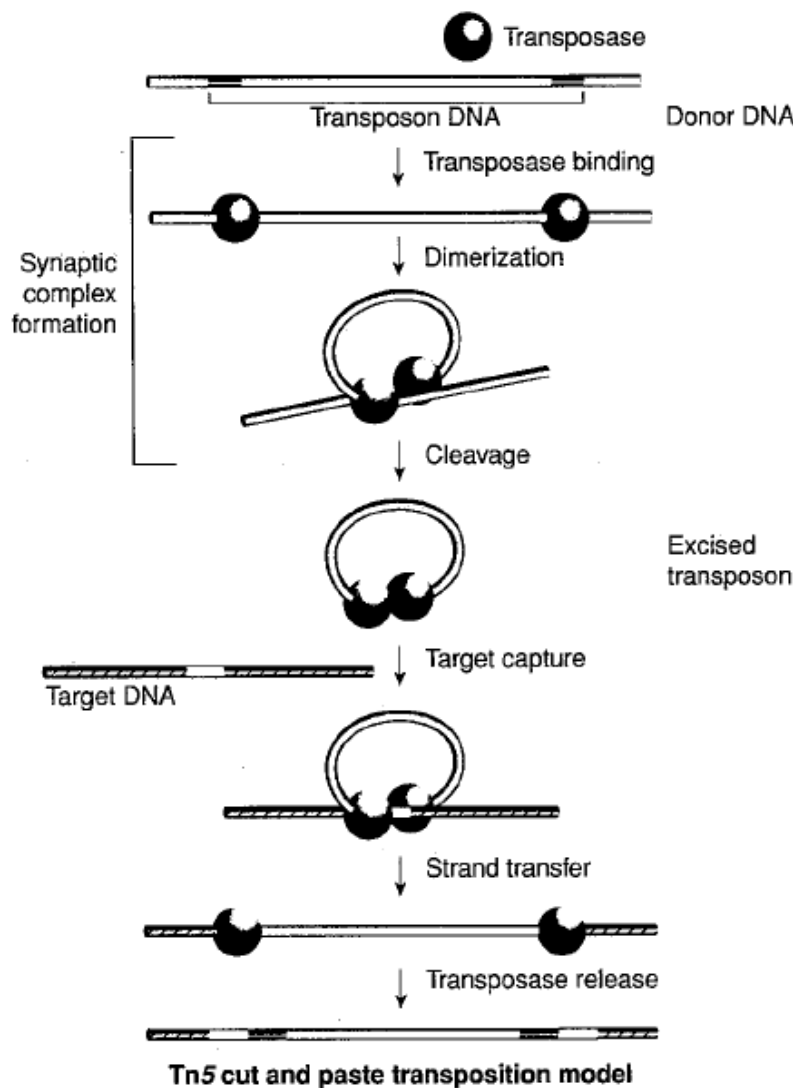


Figure 6.1: Model of Tn5 transposase-mediated transposition. The Tnp transposase binds its 19 bp target sequences followed by dimerization to form a synaptic complex. Subsequently, the synaptic complex is cleaved from the donor DNA molecule (this reaction requires Mg^{2+}) and targeted into its destination. The insertion of the transposed DNA occurs by strand transfer into 9 bp staggered cuts which results in gaps in the opposite DNA strands on each side of the insert. Host functions are thought to be required to repair these gaps, and this process thereby creates 9 bp repeats on each side of the insert. The figure is adapted from Reznikoff (2002).

Besides transformation with the pre-formed synaptic complexes, Tn5 transposition has not been tested in *S. cerevisiae*. However, excision of the full Tn5 transposon when integrated in auxotrophic markers in *S. cerevisiae* has been tested. It was shown that the Tn5 transposon could be removed from the auxotrophic marker, reverting it to its prototrophy, albeit at low frequency

[Gordenin et al., 1988 and Gordenin et al., 1991]. Because the focus of the study was the effect of the palindromic flanking sequences of Tn5 and the yeast factors contributing to its excision, it is unknown if Tn5 synaptic complexes can form *in vivo* in *S. cerevisiae*. Weil and Kunze (2000) demonstrated that another class II transposon, namely the *Ac/Ds* maize transposable element, can cause transposition in *S. cerevisiae*, in which class II transposons have not been found to naturally occur. However, the best frequency of transposition was 370 ADE prototrophic revertants per 10^7 cells, which would be rather low if the system were to be used in large scale mutagenesis. Whether the transposition frequency will be higher for Tn5 is unknown, but a hyperactive version of the Tn5 transposase (Tnp) has been isolated. The hyperactivity is due to three point mutations causing the following changes to the amino acid sequence: E54K (increases binding activity for recognition sequences), M56A (avoids synthesis of inhibitory truncated version of the transposase), and L372P (is thought to increase dimerization potential) [Goryshin and Reznikoff, 1998]. Besides the Tnp transposase from Tn5 our system needs to have the following features:

- The system needs be inducible, so transposition can be turned on when a desired cell density has been reached.
- The plasmid harboring the system needs to be able to eliminate itself to avoid any further transposition (i.e., needs to be a suicide vector).
- The plasmid needs to have minimal transcription of transposase and suicide function in *E. coli*, so the plasmid can be amplified therein.
- The proteins need to be targeted to the nucleus to effectively reach their plasmid target since the nuclear envelope of yeast stays intact during the cell cycle [Moore et al. 1998].
- The plasmid needs to contain the Tnp recognition sequences flanking the part of the plasmid which is intended for transposition into the genome of *S. cerevisiae*.
- The system needs to contain markers allowing for auxotrophic selection in yeast for plasmid-harboring cells, selection for cells where a transposition event has occurred, and selection for cells that have lost the plasmid.
- The system requires markers for selection in *E. coli* during amplification and selection for retrieving DNA fragments targeted by the transposon system.
- The system requires the ability to remove the inserted marker from the genome, enabling further rounds of mutagenesis.
- Finally the DNA fragments encompassing the system need to have a low degree of homology to the *S. cerevisiae* genome in order to limit homologous recombination.

We here describe a synthetic biology approach to create a novel random mutagenesis tool that includes all of the above listed features and therefore can effect *in vivo* mutagenesis by site-

unspecific delivery of almost any DNA sequence into the genome of *S. cerevisiae*. These characteristics allow for it to be utilized for a wide array of applications.

As the backbone plasmid for our transposon system we chose to use the pESC-URA plasmid from Stratagene (<http://people.pharmacy.purdue.edu/~thazbun/Vectors%20&%20Protocols/pESC-URA.pdf>), which contains a dual promoter system composed of the *GAL1* and *GAL10* promoters and thus allows for galactose induction of two genes. The first of these genes is the gene encoding Tnp, and the second is the gene encoding a nuclear-directed version of I-SceI [Lisby et al., 2003]. By introducing an I-SceI cleavage site in the plasmid, we converted the plasmid to a suicide vector. The cleavage site was introduced between the gene encoding the I-SceI and the promoter facilitating its transcription, hereby ensuring that I-SceI will be produced until no circular plasmid is present in the cell. Each of the genes encoding I-SceI and Tnp was spaced with an intron from *Kluyveromyces lactis* to eliminate its transcription in *E. coli* due to leakiness of the *GAL* promoters [Gibbs et al., 2004]. Furthermore, the gene encoding Tnp was C-terminally extended with the codons for the simian virus 40 nuclear localization signal (NLS) to ensure its transport into the nucleus [Lee and Melese, 1989]. The pESC-URA contains *URA3* which permits auxotrophic selection for yeast cells harboring the plasmid; it also permits counter-selection after transposition using 5-Fluoroorotic acid (5-FOA), in which case only cells having lost the plasmid will be viable.

For amplification in *E. coli* pESC-URA contains the ampicillin resistance marker. The part of the plasmid designed to transpose into the genome of *S. cerevisiae* needs to be flanked by 19 bp recognition sequences; these sequences were obtained from the pMODTM-2<MCS> plasmid from Epicentre Biotechnologies, which is designed to create DNA fragments for Tn5 transposase-mediated *in vitro* transposition. This plasmid harbors hyper-active versions of the recognition sequences known as mosaic end (ME) sequences. These sequences, in combination with the Tn5 transposase, have been found to mediate transposition of DNA fragments between 0.29 and 7 kb (upper limit is the largest fragment size tested *in vitro*) [www.epicentre.com/catalog/pmod2.htm]. To ensure efficient transposition, we kept the part of the plasmid intended for transposition below this upper limit. For selection of yeast colonies which have obtained our synthetic transposon, we included the *K. lactis ADE1* gene with its corresponding promoter and terminator. Furthermore, for retrieval of gDNA fragments which have had our transposon inserted, we included a kanamycin resistance marker. We also included loxP sites just within the ME sequences to be able to remove the majority of the transposed DNA, thereby allowing for subsequent rounds of transposon mutagenesis. The above listed DNA pieces comprise the features of the aforementioned criteria for a *S. cerevisiae in vivo* plasposon mutagenesis system. A schematic representation of our design for the system can be seen in Figure 6.2. This chapter describes the assembly of the system together with preliminary results suggesting that our transposon system is functional.

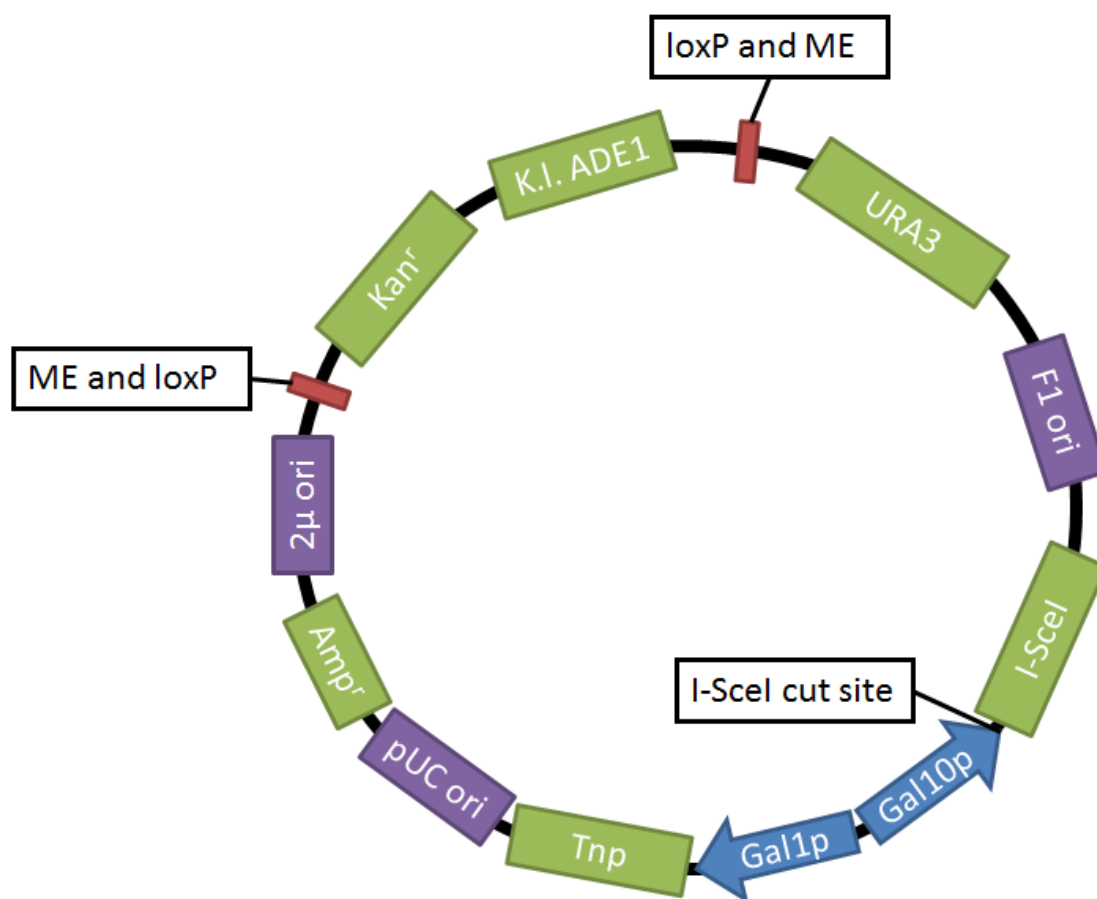


Figure 6.2: Schematic representation of the designed plasposon. This representation highlights the most important features of our design. The dual galactose inducible promoters, *Gal1p* and *Gal10p*, control the expression of the modified Tn5 transposase and the I-SceI suicide function of the plasmid, respectively. The part of the plasmid designed for transposition is flanked by the mosaic end (ME) sequence to which the transposase binds and the loxP sites to enable looping out of the majority of the transposed DNA. Furthermore, the transposable part of the plasmid contains the marker for kanamycin resistance in *E. coli* and the *ADE1* gene from *K. lactis* for auxotrophic selection in *S. cerevisiae*. The *URA3* marker allows for auxotrophic selection for cells containing the plasmid as well as counter-selection for cells having lost the plasmid. The 2μ origin of replication results in the plasmid being present with high copy number in *S. cerevisiae*, and the pUC origin of replication allows for efficient amplification in *E. coli*.

6.3 Materials and methods

6.3.1 Enzymes, chemicals and media

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich®, St. Louis, MO, and all enzymes from New England Biolabs Inc., Ipswich, MA. *E. coli* cells were grown in BD Difco™ LB broth from BD Diagnostic Systems, Sparks, MD. For selection and maintenance of plasmids, ampicillin was added to a final concentration of 100 mg/l. *S. cerevisiae* cultures that required no auxotrophic selection were grown on YPD, composed of: 10 g/l Bacto™ yeast extract, BD Diagnostic Systems; 20 g/l Bacto™ peptone, BD Diagnostic Systems; and 20 g/l dextrose. For auxotrophic selection, *S. cerevisiae* was grown on SC-medium with the appropriate nutrient(s) omitted. SC medium was composed of 6.7 g/l Yeast Nitrogen Base (YNB) without amino acids, BD Diagnostic Systems; CSM – amino acid dropout mixture (amount according to manufacturer's recommendation), Sunrise Science Products, Inc., San Diego, CA; and 20 g/l dextrose or galactose. All media components and media besides YNB were sterilized by autoclaving. YNB was sterilized by filtration.

6.3.2 Plasmid and strain construction

The components of the system were individually constructed and subsequently pieced together to form the final constructs which were used to transform *S. cerevisiae*. The plasmids used as starting points and the constructed plasmids, together with the strains applied in this study, are listed in Table 6.1. Cloning for the construction of the plasmids described below was done by treating the recipient plasmids with endonuclease(s), purification by gel-electrophoresis and recovery by gel extraction using the PureLink™ Quick Gel Extraction Kit, Life Technologies, Grand Island, NY, followed by ligation to the insert using T4 DNA ligase. The inserts were prepared either by removing it from the plasmid harboring it, *via* endonuclease digestion and gel-purification in the same way as the plasmid recipient, or amplified by PCR, purified using the QIAquick PCR Purification Kit, Qiagen, Valencia, CA, treated with endonucleases and purified again using the QIAquick PCR Purification Kit, Qiagen. PCR amplifications were performed using the Phusion® High-Fidelity DNA Polymerase. All constructs were verified by sequencing.

Table 6.1: Strains and plasmids.

Strain	Genotype	Source
<i>S. cerevisiae</i>		
F1702	<i>MATa ade1 arg4 his2 leu2 trp1 ura3</i>	Dr. Gerald Fink, Whitehead Institute, Cambridge, MA
SCX020	<i>MATa ade1 arg4 his2 leu2 trp1 - pSCX020</i>	This study
SCX021	<i>MATa ade1 arg4 his2 leu2 trp1 - pSCX021</i>	This study
SCTn009	<i>MATa arg4 his2 leu2 trp1 - pSCTn009</i>	This study
SCTn010	<i>MATa arg4 his2 leu2 trp1 - pSCTn010</i>	This study
SCTn011	<i>MATa arg4 his2 leu2 trp1 - pSCTn011</i>	This study
SCTn012	<i>MATa arg4 his2 leu2 trp1 - pSCTn012</i>	This study
<i>Escherichia coli</i>		
MAX Efficiency® DH5α™	F- φ80 <i>lacZ</i> Δ <i>M15</i> Δ(<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (rk-, mk+) <i>phoA supE44 λ- thi-1 gyrA96 relA1</i>	Life Technologies, Grand Island, NY
<i>Other</i>		
<i>Kluyveromyces lactis</i> , CLIB209 (ATCC8585)	a [k1+ k2+]	Dr. Gerald Fink, Whitehead Institute, Cambridge, MA
Plasmid name	Origin/reference	

pESC-URA	Stratagene, La Jolla, CA
pMOD2 TM	Epicentre Biotechnologies, Madison, WI
pKT127	EUROSCARF – [Sheff and Thorn, 2004]
pWJ1320	[Lisby et al., 2003]
pZE-kanFRT-tyrAaroG	Dr. Parayil Kumaran Ajikumar
MIT062	Bio Basic Inc., Markham, Ontario, Canada
pSCMOD005	This study
pSCMOD006	This study
pSCMIT	This study
pSCX020	This study
pSCX021	This study
pSCTn001	This study
pSCTn002	This study
pSCTn003	This study
pSCTn004	This study
pSCTn007	This study
pSCTn008	This study
pSCTn009	This study
pSCTn010	This study
pSCTn011	This study
pSCTn012	This study

6.3.3 Component donor plasmids

The gene encoding Tnp was purchased codon optimized for *S. cerevisiae* from Bio Basic Inc., Markham, Ontario, Canada and delivered in the MIT062 plasmid. The *K. lactis* intron KLLA0A09801g was amplified from *K. lactis*, CLIB209 gDNA by PCR using phosphorylated primers (primers 1 and 2, Table 6.2). The intron was inserted in the synthetic Tnp gene by treating MIT062 with MscI endonuclease and Antarctic phosphatase followed by T4 DNA ligase-mediated ligation. The insert direction was verified by PCR, and the plasmid with the intron inserted in the correct direction was named pSCMIT. The sequence for the intron was obtained from a list of *K. lactis* introns published at http://genome.jouy.inra.fr/genosplicing/Seq/KLLA_introns.tfa.

The transposable cassette was constructed based on the pMOD2TM plasmid from Epicentre Biotechnologies, Madison, WI. The *K. lactis* *ADE1* gene was amplified from CLIB209 gDNA by PCR using primers 3 and 4; these primers contain adaptamers with the loxP sites flanked by SalI and BamHI restriction sites, respectively. The amplified DNA fragment and pMOD2TM were treated with SalI and BamHI, purified, and ligated together using T4 DNA ligase. The constructed plasmid was named pSCMOD005. The *KanR* gene encoding kanamycin resistance was amplified from the pZE-kanFRT-tyrAaroG plasmid by PCR using primers 5 and 6. The amplified *KanR* fragment and pSCMOD005 were treated with MluI and purified. *KanR* was ligated into pSCMOD005 using T4 DNA ligase to create pSCMOD006. This plasmid serves as donor DNA for the transposable part of the final plasmid.

6.3.4 Plasmids for verification of nuclear localization of the Tnp transposase

The main component of the system, the transposase, has to be located in the nucleus to be able to catalyze the transposition of part of the plasmid into the genome. We therefore chose to C-terminally extend the gene encoding it with a 21 bp sequence encoding the simian virus 40 NLS composed of 7 amino acids (Pro-Lys-Lys-Lys-Arg-Lys-Val) [Kalderon et al., 1984]. For verification that the NLS-fused transposase does indeed localize to the nucleus, we constructed a plasmid harboring *Tnp-NLS* fused to the 3' of the gene encoding the yeast enhanced GFP (*yeGFP*). This was done by first amplifying *yeGFP* from pKT127 using primers 7 and 8. The adaptamer of primer 8 contained a DNA stretch encoding a 16 amino acid linker, based on the linker tested by Chang et al. (2005). The *yeGFP* fragment was inserted in the pESC-URA plasmid in between the BamHI and NheI endonuclease cleavage sites. The resulting plasmid was named pSCX020. The *Tnp* gene, spaced with the *K. lactis* intron, was amplified from pSCTn007 (see below) using primers 9 and 10 and inserted in the NheI site of pSCX020 to create pSCX021. The insert directionality was verified by PCR. pSCX020 and pSCX021 were used to test whether the GFP transposase co-localizes with the nucleus, which we stained with 4',6-Diamidino-2-phenylindole dihydrochloride (DAPI).

6.3.5 Complete system

The construction of the final plasmid containing all the aforementioned components, and the control plasmids, were based on the pESC-URA. pESC-URA was modified by inserting I-SceI in between the EcoRI and PacI restriction sites. I-SceI was amplified by PCR from the pWJ1320 plasmid using primers 11 + 13 and 12 + 13. The fragments amplified with primers 11 and 13 contained the I-SceI cut-site upstream of the open reading frame. Thus the insertion of the two fragments created two plasmids, one with the I-SceI cut-site and one without it, named pSCTn001 and pSCTn002, respectively. These two plasmids were both modified by inserting the KLLA0A09801g intron in the SfoI site, to create pSCTn003 and pSCTn004, respectively. The intron was amplified by PCR from *K. lactis*, CLIB209 gDNA, using primer 1 and 2. pSCTn003 and pSCTn004 were modified by inserting the intron-modified *Tnp* gene in between the BamHI and SacII restriction sites to create pSCTn007 and pSCTn008. The *Tnp* fragment was amplified by PCR from pSCMIT applying primers 14 and 15. Primer 15 contained an adaptamer containing a DNA sequence encoding the simian virus NLS and a new stop codon, hereby modifying the *Tnp* gene further before it was inserted in the two plasmids. The cassette destined for transposition was amplified by PCR from pSCMOD006 using primers 16 and 17. The forward and reverse primers contained adaptamers containing sequences identical to the sequences of each side of the MfeI restriction site, respectively, located in the pSCTn003, pSCTn004, pSCTn007, and pSCTn008. The amplified cassette was then inserted in pSCTn003, pSCTn004, pSCTn007, and pSCTn008 treated with MfeI by gap-repair in *S. cerevisiae* to create pSCTn009, pSCTn010, pSCTn011, and pSCTn012, respectively.

Table 6.2: Primers. Restriction sites used for cloning are underlined.

#	Name	Sequence (5'-3')	5'-Modification
1	KL09801g-intron-for	GTATGTAAACCATGTATGCGTTTACTATTA TT	Phosphorylated
2	KL09801g-intron-rev	CTGGGGAGATAAGTGGGATTTC	Phosphorylated
3	K.l.ADE1-for-loxP-SalI	GACTTAG <u>TCGAC</u> Cataactcgtataatgtatgctatacgaagt atGTAACGCGTGTATAAATATATCTCTG ¹	None
4	K.l.ADE1-rev-loxP-BamHI	TATGAC <u>GGATCC</u> Cataactcgtatagcatacattatacgaagt atCAAGTCGGTGAAATGCTAGAACA ¹	None
5	KanR-for-MluI	GACTTA <u>ACGCGT</u> AGAGCGCTTTTGAAGCTC ACGC	None
6	KanR-rev-MluI	TATGAC <u>ACGCGT</u> AAGATCCCCTTATTAGAA GAACTCGTC	None
7	GFP-for-BamHI	GACTTAGGATCCATGTCTAAAGGTGAAGAA TTATTCATCTGG	None
8	GFP-rev-linker-NheI	TATGACGCTAGCtgcagcttcaccagcaccagctgcagatg cagcagaaccagaagtTTTGTACAATTCATCCATAC CATGGGTAATA ²	None
9	Tn5-for-NheI	GACTTAGCTAGCATGATAACATCTGCTTTA CATAGAGCCG	None
10	Tn5-rev-NheI	TATGACGCTAGCTTAACCAACTTTTCTCTTT TTCTTTGGTATCTTTAT	None
11	I-SceI-for-EcoRI-CS	GACTTAGAATTCttacgctaggataacagggtaatatagcg ATGGGATCAAGATCGCCAAAAAAGAAG ³	None
12	I-SceI-for-EcoRI	GACTTAGAATTCATGGGATCAAGATCGCCA AAAAAGAAG	None
13	I-SceI-rev-PacI	TATGACTTAATTAATTATTTTCAGGAAAGTTT CGGAGGAGATAG	None
14	Tn5-for-BamHI	GACTTAGGATCCATGATAACATCTGCTTTA	None

		CATAGAGCC	
15	Tn5-rev-SV40-SacII	TATGACCCGCGGTTAACCaaactttctcttttctttgTA TCTTTATTCCTTGAGCCATCAAATCC ⁴	None
16	pMOD006-for	ttatccattccatgcggggtatcgatgcttcctcagcactacccttag ctgttctatatgtgccactcctGCGCGTTGGCCGATTCA TTAAGA ⁵	None
17	pMOD006-rev	atgtcatgataaatgggttcttagtatgatccaatatcaaaggaaatgat agcattgaaggatgagactaatcCGGGCCTCTTCGCTAT TACGA ⁵	None

Features included in the adaptamers marked with lowercase letters:

¹ loxP sites.

² Linker sequence.

³ I-SceI endonuclease cleavage site.

⁴ Simian virus 40 NLS.

⁵ Sequences for homologous recombination with pSCTn003, pscTn004, pSCTn007 and pSCTn008.

6.3.6 Strains

All transformations of *S. cerevisiae* were carried out using the lithium acetate/polyethylene glycol method according to Gietz and Woods [2002]. Transformants were selected on SC medium with the appropriate nutrient(s) omitted. The constructed strains were verified by extraction of total DNA using the Wizard® Genomic DNA Purification Kit, Promega, Madison, WI, followed by PCR verification that the plasmid DNA was present therein. All the constructed strains were based on the F1702 strain. Each transformant was named after the plasmid used to transform it by removing the initial ‘p’ of the plasmid name, e.g. the resultant strain of F1702 being transformed with pSCTn011 was named SCTn011.

6.3.7 Evaluating nuclear localization

The potential for nuclear localization of the modified *Tnp* gene was tested *in silico* before the actual construct was made. For this purpose the NLS mapper available at http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi was applied [Kosugi et al., 2009]. To test if the modified Tnp transposase *in vivo* localizes to the nucleus of *S. cerevisiae*, the SCX020 and

SCX021 strains were grown in SC-URA (glucose) overnight. The cells were harvested by centrifugation, washed with 5 ml sterile dH₂O and resuspended in 5 ml sterile dH₂O. The OD₆₀₀ of each of the cell suspensions was measured, and the cells were inoculated into SC-URA (galactose) and incubated for 5 h at 30 °C. After the first four hours of the incubation the tubes were covered with aluminum foil and 5 µl DAPI (1 mg/ml) was added. After the last hour of incubation the cells were inspected by fluorescence microscopy; 5 µl of each cell suspension was transferred to a microscope slide and examined using the Nikon Eclipse TE200-S microscope, Nikon Inc., Melville, NY, equipped with 100x objective, GFP band-pass filter (GFP-B: EX480/40, DM505, BA535/50) and DAPI band-pass filter (UV-2E/C: EX340-380, DM400, BA435-485). Fluorescent images were acquired using the following exposure times: 60 msec for bright field, 3 sec for DAPI, 150 msec for GFP images of SCX020, and 6 sec for GFP images of SCX021. The images were processed using the SPOT AdvanceTM software, SPOTTM, Imaging Solutions, a division of Diagnostic Instruments, Inc., Sterling Heights, MI.

6.3.8 Plasposon system testing

All of the cells from each of the F1702 transformants with pSCTn009, pSCTn010, pSCTn011, and pSCTn012 were scraped off the SC-URA (glucose) plates on which they had been selected using 1 ml of cold water and a Drigalski spatula. The cells were transferred to 250 ml shake flasks each containing 50 ml SC-ADE (galactose) medium. The initial OD₆₀₀ were measured, and the shake flasks incubated at 30 °C with 250 rpm orbital shaking. The growth of the four cultures was monitored by measuring the OD₆₀₀.

6.4 Results and discussion

6.4.1 Nuclear localization

The designed *Tnp* gene encoding the transposase engineered for nuclear localization was tested *in silico* using the online available NLS mapper to give an indication of its nuclear localization potential. The best NLS identified within the designed protein sequence was assigned a score of 10.5, whereas the best NLS identified within the protein without the NLS was assigned a score of 3. The NLS mapper was experimentally tested on different versions of the GUS-GFP fusion proteins, and it was found that these fusion proteins with a score higher than eight were exclusively located in the nucleus, whereas proteins with a score between three and five were located both in the nucleus and the cytoplasm [Kosugi et al., 2009].

After constructing the gene encoding the nuclear directed Tnp, it was tested for its ability to localize to the nucleus by tagging it with yeGFP and using fluorescence microscopy (Figure 6.3). Comparing the two strains, the result of the DAPI stain is very similar, whereas the GFP images are very different. GFP expressed in SCX020 is not linked to the nuclear Tnp transposase and can therefore be seen throughout the cell. In SCX021, where GFP has been linked to the transposase, the GFP signal can be seen to co-localize with the DAPI signal. In conclusion, we discovered that the yeGFP-tagged protein localizes to the nucleus stained with DAPI (Figure 6.3). The yeGFP expressed by SCX020 was located throughout the cell. These results show that our engineered Tnp localizes to the nucleus.

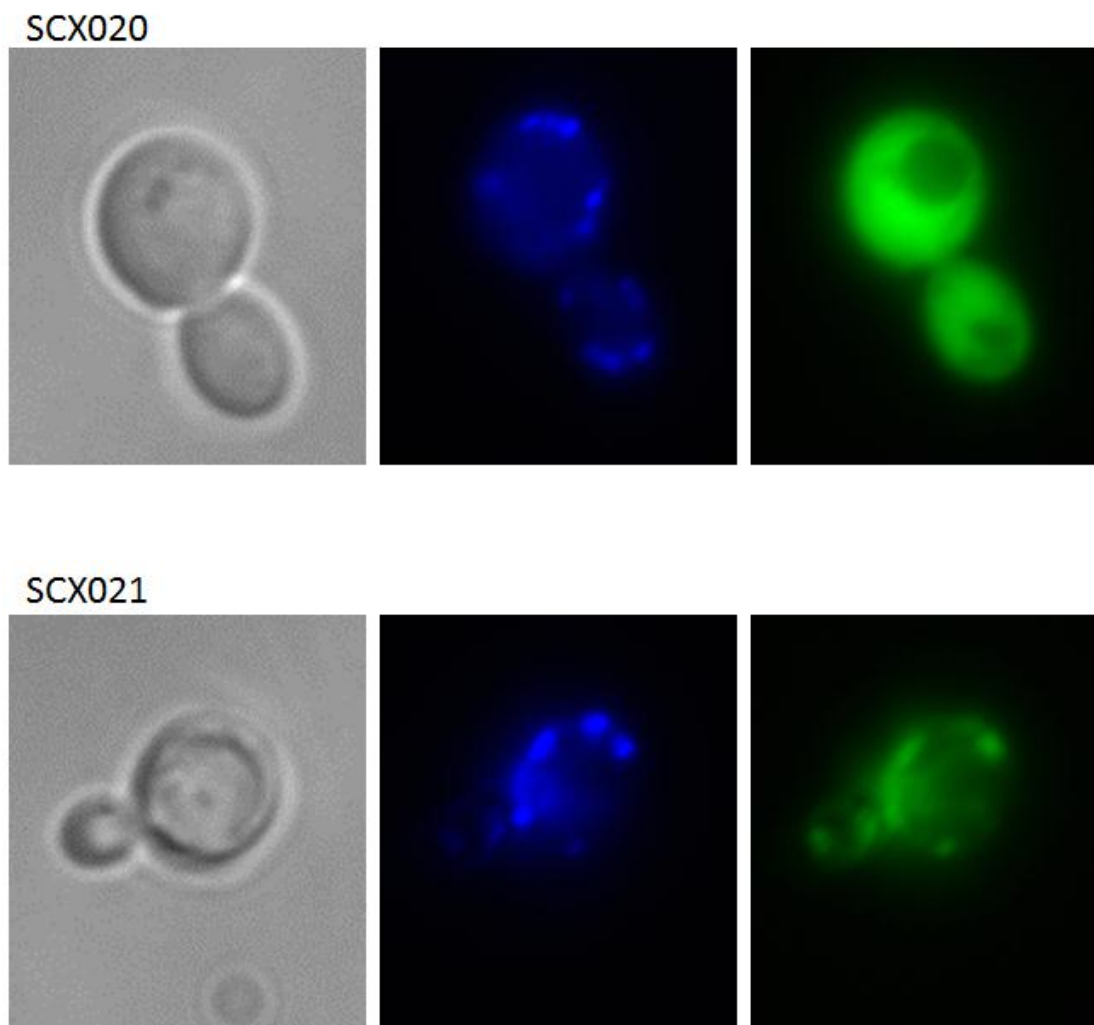


Figure 6.3: Fluorescence microscopy images. All images were acquired with 100X objective. The first and second rows correspond to images of SCX020 and SCX021, respectively. The first column contains the bright field images; the second column is the result from DAPI staining; and the third column is the result of GFP fluorescence.

The exposure times used to acquire the bright field and DAPI images were the same for both the SCX020 and SCX021. However, the exposure time used to acquire the GFP image of SCX021 cells was 40 times longer than for the SCX020. The reason for this is probably caused by lower expression of the GFP-tagged transposase. The lower expression can be caused by the position of intron within the gene. Yoshimatsu and Nagawa (1994) found that the expression level of proteins in *S. cerevisiae* could be changed by inserting an intron. The closer the intron was located to the start codon of the gene encoding the protein, the higher the expression. By tagging our transposase N-terminally we shifted the relative position of the engineered intron away from the start codon, hereby potentially causing lower expression of the entire fusion protein. We could have avoided the lower expression by tagging the transposase C-terminally, but this would have positioned the NLS between GFP and Tnp, creating the possibility that the fusion protein

could be degraded by proteases or in another way fragmented such that the NLS would only direct the GFP part of the protein to the nucleus. We chose not to construct and test the C-terminally tagged Tnp since any result obtained thereof would be inconclusive due to the increased risk of having false positives.

6.4.2 Plasposon system

The plasmid pSCTn011 encompasses our design for a plasposon system based on the bacterial Tn5 transposase. Besides this plasmid we constructed 3 control plasmids, namely pSCTn009, pSCTn010, and pSCTn012. Maps for the four plasmids can be seen in Figure 6.4. pSCTn009 is designed to test the effect of the suicide function engineered into the plasmid without the effect of the transposase; pSCTn010 is designed as the negative control of the system, having neither suicide function nor transposase; and pSCTn012 is designed to test the effect of the transposase without the suicide function. When *S. cerevisiae* cells harboring the plasmid containing the suicide function are induced on galactose, all plasmids should be linearized and unable to propagate or to be passed on to subsequent generations of cells; therefore, only cells capable of mobilizing the *K. lactis ADE1* gene to a locus outside the plasmid should be viable. On the other hand, cells harboring the plasmids without the suicide function should be unaffected by galactose induction unless the transposase has a detrimental effect on the cell growth.

To test how the cells transformed with the four plasmids would be affected by induction on galactose, for each individual transformation all cells were inoculated into 50 ml SC-ADE (galactose) and the growth was monitored. The resulting growth curves can be seen in Figure 6.5. It can be seen that SCTn011 grows much faster than SCTn009 and almost as well as the negative control SCTn010 and SCTn012, which indicates that the transposase has a positive effect on growth by alleviating some of the detrimental effects caused by the suicide mechanism. This result indicates that the transposase is active and can mobilize the cassette containing the *K. lactis ADE1* gene from the plasmid to the genome. By comparing SCTn10 and SCTn012 it can be seen that the transposase by itself has very little effect on growth, whereas comparing SCTn009 and SCTn010 shows that there is a large impact on growth by the suicide mechanism. Late in the course of the growth experiment SCTn009 starts to grow even though it should be incapable of growth. This can be caused by linearized plasmids being integrated into the genome by non-homologous end joining or any other mechanism thwarting the auxotrophic selection. Genomic integration of the linearized plasmid will not be a problem when the cells are finally selected on SC-ADE+5-FOA since these hereby will be eliminated. A schematic representation of the method of application and possible outcomes of applying the system can be seen in Figure 6.6.

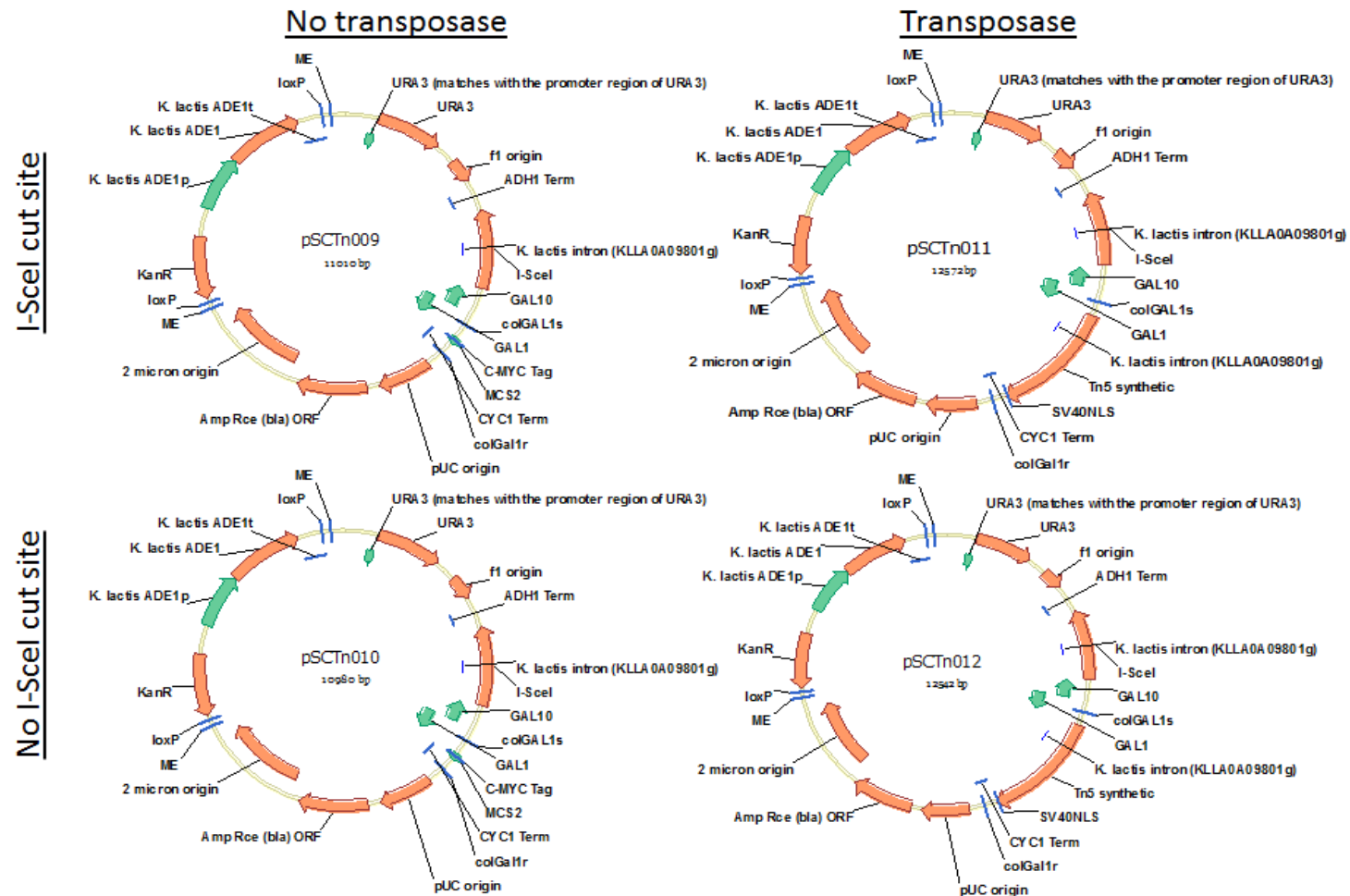


Figure 6.4: Plasmid maps. Plasmid maps for the pSCTn009, pSCTn010, pSCTn011, and pSCTn012 plasmids used to test the functionality of the two main components of the plasposon system, namely the transposase and the suicide mechanism.

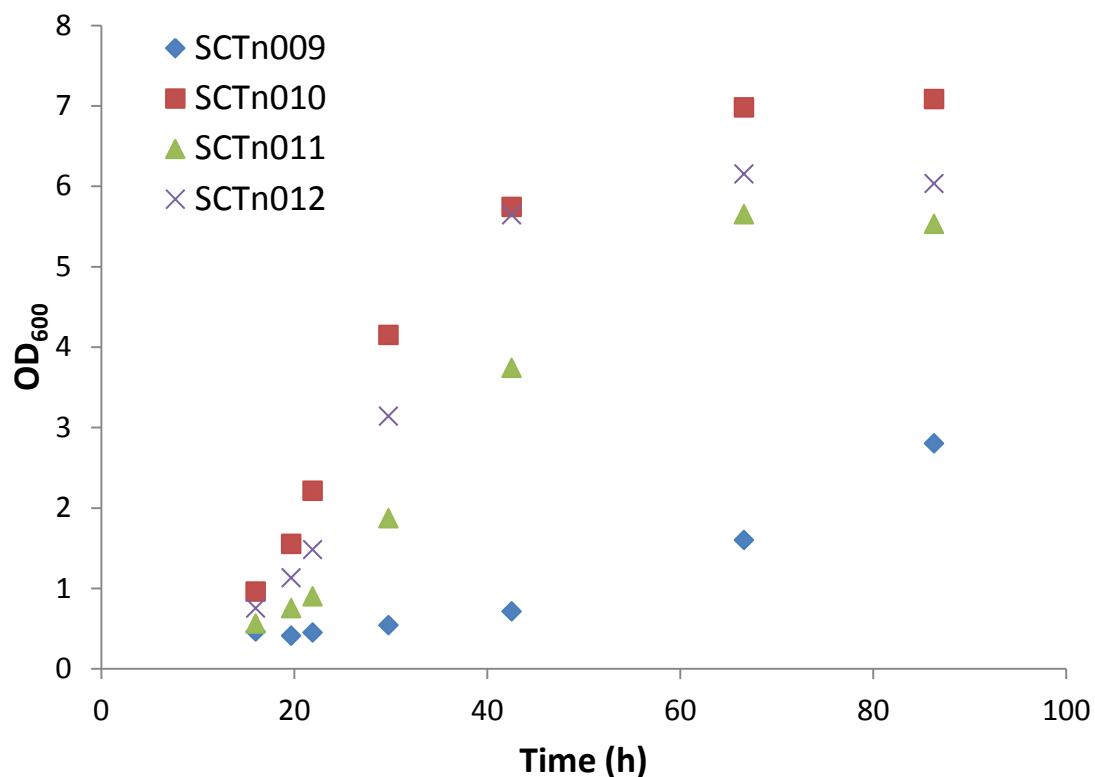


Figure 6.5: Growth curves for SCTn009, SCTn010, SCTn011, and SCTn012. The cells were scraped off the plates used for selection of the transformants harboring the plasmid and inoculated in 250 ml shake flask containing 50 ml SC-ADE (galactose) medium. The growth of the strains was monitored by measuring the OD₆₀₀ and plotted as a function of sampling time.

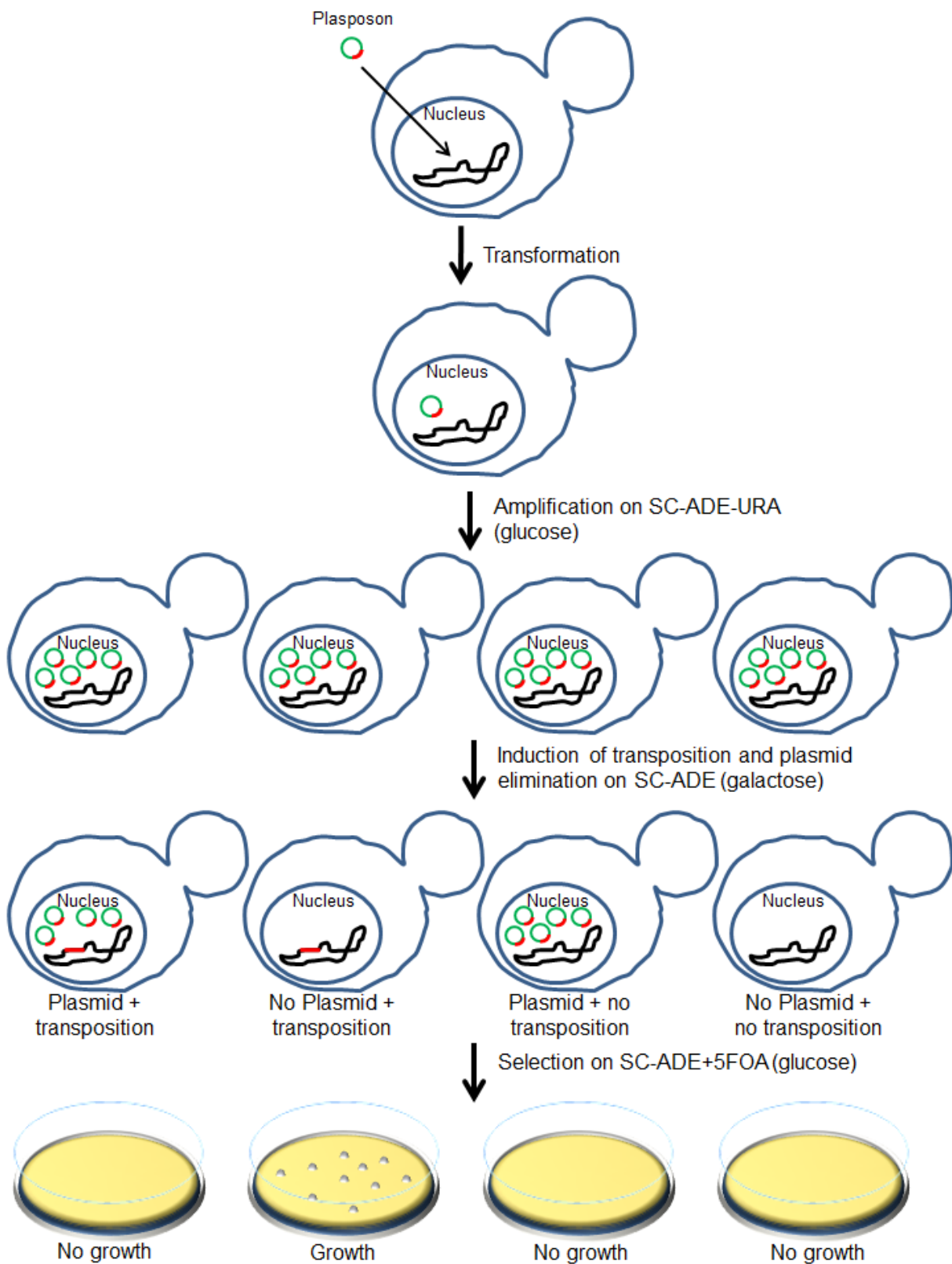


Figure 6.6: Schematic representation of the application of the plasposon system for *in vivo* mutagenesis of *S. cerevisiae*. The plasposon is first transformed into the *S. cerevisiae* strain chosen as starting point for the mutagenesis, followed by amplification of the cells and propagation of the plasposon on SC-ADE-URE (glucose) medium. Subsequently the transposition is induced by shifting the cells to SC-ADE (galactose) medium which only selects for the transposable part of the plasmid. Four different outcomes of this induction can occur: transposition with the cells maintaining the plasmid, transposition with plasmid loss, no transposition with the cells maintaining the plasmid, and no transposition with plasmid loss. If these cells are transferred to SC-ADE+5-FOA medium, only the desired outcome, where cells experiencing transposition and plasmid loss, will be viable. These random insert mutants can then be screened for a secondary phenotype for which one wants to elucidate the causative genotype.

6.4.3 Further experiments

The results presented in this chapter are preliminary and only suggest that the system is functional. To develop this system and prove that it works, further experiments are required. The first experiment required will be to repeat the growth experiment previously performed with all cells from the transformation and carry it out with cells picked from single colonies. Secondly, an experiment proving that the introns inserted in the two genes are efficiently spliced out should be performed. This could be done by growing the SCTn012 cells on SC-ADE-URA (galactose) from which cDNA could be purified. The purified cDNA could then be probed using PCR with primers annealing on each site of the intron. If only one band occurs, the size of the band will indicate whether the intron is efficiently spliced out, whereas if two bands occur, the ratio between the bands will indicate the proportion of mRNA molecules without the intron compared to mRNA molecules with the intron. Finally, a gDNA library in an *E. coli* plasmid should be created from gDNA isolated from strains found, by PCR, not to harbor the plasmid backbone, but only the sequence designed to be transposed. The kanamycin resistance marker should allow for *E. coli* cells transformed with the library to be selected, so only cells transformed with a plasmid containing a transposon fragment will be viable. The plasmids isolated from these transformants should be sent for sequencing to test whether the insert is flanked by 9 bp repeats. If the sequencing results confirm that the insert is indeed flanked by 9 bp repeats, it proves that our plasposon system is functional. If the system is found to be functional, an experiment to test the frequency of transposition could be carried out to benchmark our system against other transposition systems. All together these experiments could prove the functionality of our system and allow for the system to be implemented as an efficient tool for random DNA delivery into the genome of *S. cerevisiae*.

Since the loci of genomic integration have been found to affect the expression level of the inserted genes [Flagfeldt et al. 2009], our system could be used to probe the entire genome of *S. cerevisiae* for ‘expression hotspots.’ The system could be set up to deliver a GFP expression cassette into the genome, and the resulting mutants could be selected by FACS for those exhibiting the highest degree of fluorescence. Such an experiment would demonstrate the applicability of our system.

6.5 References

Chang H-C, Kaiser CM, Hartl FU, Barral JM: ***De novo* folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria.** J. Mol. Biol. 2005, 353:397-409.

Curcio MJ, Garfinkel DJ: **Single-step selection for TyI element retrotransposition.** Proc. Natl. Acad. Sci. USA. 1991, 88:936-940.

Dennis JJ, Zylstra GJ: **Plasposon: modular self-cloning minitransposon derivatives for rapid genetic analysis of Gram-negative bacterial genomes.** Appl. Environ. Microb. 1998, 64(7):2710-2715.

Devine SE, Boeke JD: **Integration of the yeast retrotransposon TyI is targeted to regions upstream of genes transcribed by RNA polymerase III genes.** Genes Dev. 1996, 10:620-633.

Flagfeldt DB, Siewers V, Huang L, Nielsen J: **Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*.** Yeast. 2009, 26:545-551.

Gibbs MD, Reeves RA, Sunna A, Bergquist PL: **A yeast intron as a translational terminator in a plasmid shuttle vector.** FEMS Yeast Res. 2004, 4:573-577.

Gietz RD, Woods RA: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.** Method. Enzymol. 2003, 350:87-96.

Gordenin DA, Trofimova MV, Shaburova ON, Pavlov YI, Chernoff YO, Chekuolene YV, Proscyavichus YY, Sasnauskas KV, Janulaitis AA: **Precise excision of bacterial transposon Tn5 in yeast.** Mol. Gen. Genet. 1988, 213:388-393.

Gordenin DA, Proscyavichus YY, Malkova AL, Trofimova MV, Peterzen A: **Yeast mutants with increased bacterial transposon Tn5 excision.** Yeast. 1991, 7:37-50.

Goryshin IY, Reznikoff WS: **Tn5 *in vitro* transposition.** J. Biol. Chem. 1998, 273(13):7367-7374.

Goryshin IY, Jendrisak J, Hoffman LM, Meis R, Reznikoff WS: **Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes.** Nat. Biotechnol. 2000, 18:97-100.

Kalderon D, Roberts BL, Richardson WD, Smith AE: **A short amino acid sequence able to specify nuclear localization.** Cell. 1984, 39:499-509.

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF: **Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence.** Genome Res. 1998, 8:464-478.

Kosugi S, Hasebe M, Tomita M, Yanagawa H: **Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs.** Proc. Natl. Acad. Sci. USA. 2009, 106(25):10171-10176.

Lee W-C, Melese T: **Identification and characterization of a nuclear localization sequence-binding protein in yeast.** Proc. Natl. Acad. Sci. USA. 1989, 86:8808-8812.

Lisby M, Mortensen UH, Rothstein R: **Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre.** Nat Cell Biol. 2003, 5(6):572-577.

McClintock B: **The origin and behavior of mutable loci in maize.** Genetics. 1950, 36:344-355.

Natsoulis G, Thomas W, Roghmann M-C, Winston F, Boeke JD: **TyI transposition in *Saccharomyces cerevisiae* is non-random.** Genetics. 1989, 123: 269-279.

Reznikoff WS: **Tn5 transposition**. IN: Chapter 18, Mobile DNAIL. Edited by Craigh NL. ASM Press 2002, Washington, D.C.

Seifert HS, Chen EY, So M, Heffron F: **Shuttle mutagenesis: a method of transposon mutagenesis for *Saccharomyces cerevisiae***. Proc. Natl. Acad. Sci. USA. 1986, 83:735-739.

Smith C: **Transposon: cut-and-paste gene delivery**. Nat. Methods. 2007, 4:183-186.

Sheff MA, Thorn KS: **Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae***. Yeast. 2004, 21:661-670.

Yoshimatsu T, Nagawa F: **Effect of artificially inserted intron on gene expression in *Saccharomyces cerevisiae***. DNA Cell Biol. 1994, 13(1):51-58.

Vizváryová M, Valkov D: **Transposons – the useful genetic tools**. Biologia. 2004, 59(3):309-318.

Weil CF, Kunze R: **Transposition of maize *Ac/Ds* transposable elements in the yeast *Saccharomyces cerevisiae***. Nat. Genet. 2000, 26:187-190.

Chapter 7 – Engineering *Escherichia coli* for Production of Squalene: Utilization of the *ERG9* Gene from *Saccharomyces cerevisiae*

Simon Carlsen^{*†}, Haoran Zhang[†], Marjan De Mey^{†‡}, Michael Lyng Nielsen^{*·¹}, Anna Eliasson Lantz^{*}, Morten C. Kielland-Brandt^{*}, Gregory Stephanopoulos^{†§}

[§]Corresponding author

^{*} Department of Systems Biology, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

[†] Department of Chemical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA.

[‡] Centre of Expertise – Industrial Biotechnology and Biocatalysis, Ghent University, Coupure links 653, 9000 Ghent, Belgium

¹ Present address: Novozymes A/S, Krogshøjvej 36, DK-2880 Bagsvaerd, Denmark.

Manuscript intended for submission to Biotechnology and Bioengineering

7.1 Abstract

The triterpene compound squalene is of great commercial interest due to its unique properties which makes it amenable for application in a wide variety of products. Squalene has traditionally been isolated from shark liver oil. Environmental concerns and lack of success in obtaining this commodity chemically have stimulated the search for alternatives such as plant extraction and production by microbial fermentation. Especially the latter has gained a lot of attention as this may yield a sustainable alternative resource of squalene. Until recently only microorganisms that naturally produce squalene, such as yeasts, have been applied in this endeavor. However, the use of a native host might not be the best choice as production organism due to its intrinsic metabolic regulation. To by-pass this native regulation we decided to engineer *Escherichia coli* for the production of squalene by expression of squalene synthase encoded by the *ERG9* gene from *Saccharomyces cerevisiae* together with the overexpression of the native *E. coli* farnesyl pyrophosphate (FPP)-synthase, encoded by the *ispA* gene. We here demonstrate that both the full-length and the 24 amino acids C-terminally truncated *ERG9* gene can be used for heterologous production of squalene. Furthermore, the results indicate that squalene exhibit an inhibitory effect on *E. coli* cell growth. The plasmid mediating squalene production was introduced into three strains with different levels of methylerythritol 4-phosphate pathway expression, of which the best strain produced 3.04 mg/l/OD₆₀₀.

Keywords: *Escherichia coli*, squalene, metabolic engineering, *ERG9*, squalene synthase.

The triterpene molecule squalene was first identified in 1916 by M. Tsujimoto, who isolated this hydrocarbon, with the empirical formula $C_{30}H_{50}$, from shark liver oil (Tsujimoto 1920). One of the main applications of squalene is as an emollient in cosmetics and pharmaceuticals. Furthermore squalene is used in industrial products such as magnetic tape and low-temperature lubricants (Bhattacharjee et al., 1993). Squalene has also been found to have beneficial effects in the protection against cancer, coronary heart disease, aging, and bacterial infections. For a review on the healthcare aspects of squalene see Reddy and Couvreur (2009); for a more general review on squalene, see Spanova and Daum (2011). Problems with high mercury concentrations in shark liver, environmental/marine protection concerns, and animal rights groups (e.g. PETA and Oceana) urging consumers to boycott products containing squalene derived from animals, have prompted the search for alternative production processes of squalene (Bondioli et al., 1993; Bhattacharjee et al., 2001; He et al., 2002).

Besides its occurrence in shark liver oil, a wide range of organisms such as bacteria, fungi, algae, plants and mammals have been found to synthesize squalene. Squalene belongs to the group of secondary metabolites known as isoprenoids or terpenoids. It is biosynthesized from isopentenyl pyrophosphate (IPP), which is the universal precursor of all isoprenoids. In nature, IPP is synthesized from central carbon metabolites via two pathways: the mevalonate (MVA)-pathway and the methylerythritol phosphate (MEP)-pathway. IPP is then isomerized to form dimethylallyl pyrophosphate (DMAPP). Subsequently DMAPP reacts with IPP and another molecule of DMAPP to form farnesyl pyrophosphate (FPP), the precursor of sesquiterpenoids such as squalene (Ajikumar et al., 2008). This reaction is catalyzed by the enzyme FPP synthase (EC 2.5.1.10), which in *E. coli* is encoded by the *ispA* gene (Fujisaki et al., 1990). Next, two molecules of FPP are condensed head-to-head to form squalene. This reaction occurs *via* the intermediate pre-squalene, which is finally reduced by NADPH to form squalene. This final step is catalyzed by the enzyme squalene synthase (EC 2.5.1.21) (Agnew and Popják, 1977). The first squalene synthetase to be isolated and characterized was from the microsomes of *Saccharomyces cerevisiae* (Lynen et al., 1958), later discovered simultaneously by Fegueur et al. (1991) and Jennings et al. (1991) to be encoded by the *ERG9* gene from *S. cerevisiae*. Since squalene synthase catalyzes the first committed step in sterol biosynthesis, it is an obvious target for the treatment of hypercholesterolemia, and the initial efforts were therefore directed towards obtaining purified soluble squalene synthase for *in vitro* characterization, which could aid the identification of novel therapeutics. The squalene synthase was found to be anchored to microsomal membranes in part by a C-terminal stretch of 24 hydrophobic amino acids. By removing the codons for these amino acids and expressing the truncated gene in *E. coli*, it has been possible to produce a soluble version of the enzyme which has been thoroughly characterized *in vitro* (LoGrasso et al., 1993; Zhang et al., 1993). Both the full-length and the truncated *ERG9* have been expressed in *E. coli* for the purpose of purification and characterization. However, neither of the two has so far been tested for the ability to mediate *in vivo* synthesis of squalene when heterologously expressed.

To investigate whether the *ERG9* gene from *S. cerevisiae* can be used to engineer *E. coli*, which does not naturally, synthesizes squalene, in order to produce squalene, we constructed two plasmids, p10T5T10-ERG9 and p10T5T10-ERG9t, harboring synthetic operons for expressing the full-length and truncated *ERG9* gene, respectively, in conjunction with the *E. coli ispA* gene. Each of the two plasmids were used independently to transform the *E. coli* strains K-12 MG1655, K-12 MG1655 TRC-MEP, and K-12 MG1655 T7-MEP, previously used by Ajikumar et al. (2010) for high-titer production of taxadiene. The last two of these strains have been engineered to express a chromosomally integrated operon containing the *dxs*, *idi*, *ispD* and *ispF* genes under the control of the TRC- and T7-promoter, respectively. All plasmids and strains used in this study are listed in Table 7.1.

Table 7.1: Strains and plasmids.

Strain	Genotype	Source
<i>Escherichia coli</i>		
K-12, MG1655	$\Delta recA \Delta endA DE3$	Dr. Parayil Kumaran Ajikumar*
K-12, MG1655 TRC-MEP	$\Delta recA \Delta endA DE3$	Dr. Parayil Kumaran Ajikumar*
K-12, MG1655 T7-MEP	$\Delta recA \Delta endA DE3$	Dr. Parayil Kumaran Ajikumar*
MAX Efficiency® DH5 α ™	<i>F- ϕ80lacZAM15 Δ(lacZYA- argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ- thi-1 gyrA96 relA1</i>	Life Technologies, Grand Island, NY
<i>Saccharomyces cerevisiae</i>		
F1702	<i>MATa ade1 arg4 his2 leu2 trp1 ura3</i>	Dr. Gerald Fink, Whitehead Institute, Cambridge, MA
Plasmid name	Origin/reference	
p10T5T10	Dr. Parayil Kumaran Ajikumar	
p10T5T10-ERG9	This study	
p10T5T10-ERG9t	This study	

* (Ajikumar et al., 2010)

To evaluate the constructed strains they were inoculated at OD₆₀₀ = 0.01 from pre-cultures, induced with 0.1 mM IPTG at OD₆₀₀ = 0.1 and grown for 3 days at 20 °C followed by extraction into acetone. The cell extracts were analyzed by GC-FID and GC-MS. By comparison to an authentic squalene standard it was verified that the six strains were able to produce squalene, cf. Figure 7.1 and Figure 7.2. These results show that the major compound extracted has the same retention time and MS-spectrum as the squalene standard, hereby verifying that the strains are capable of producing squalene.

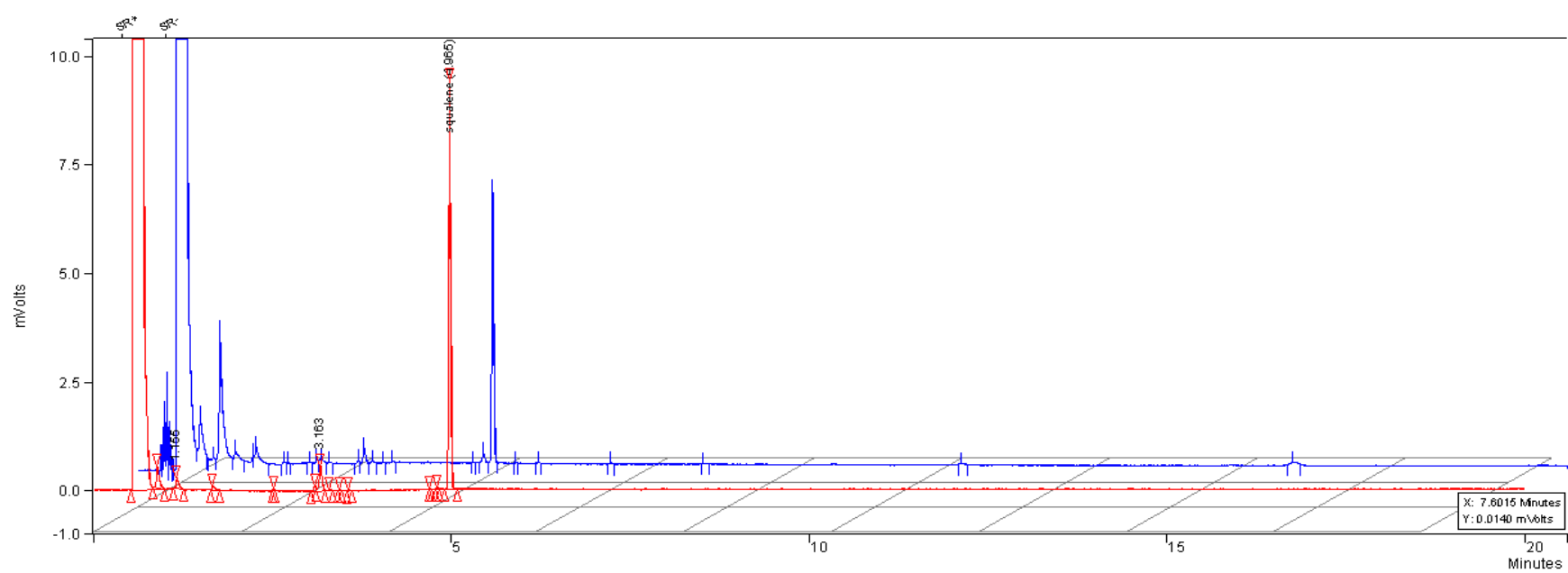


Figure 7.1: GC-FID chromatogram of squalene. Red graph – squalene standard and blue graph – *E. coli* cell extract.

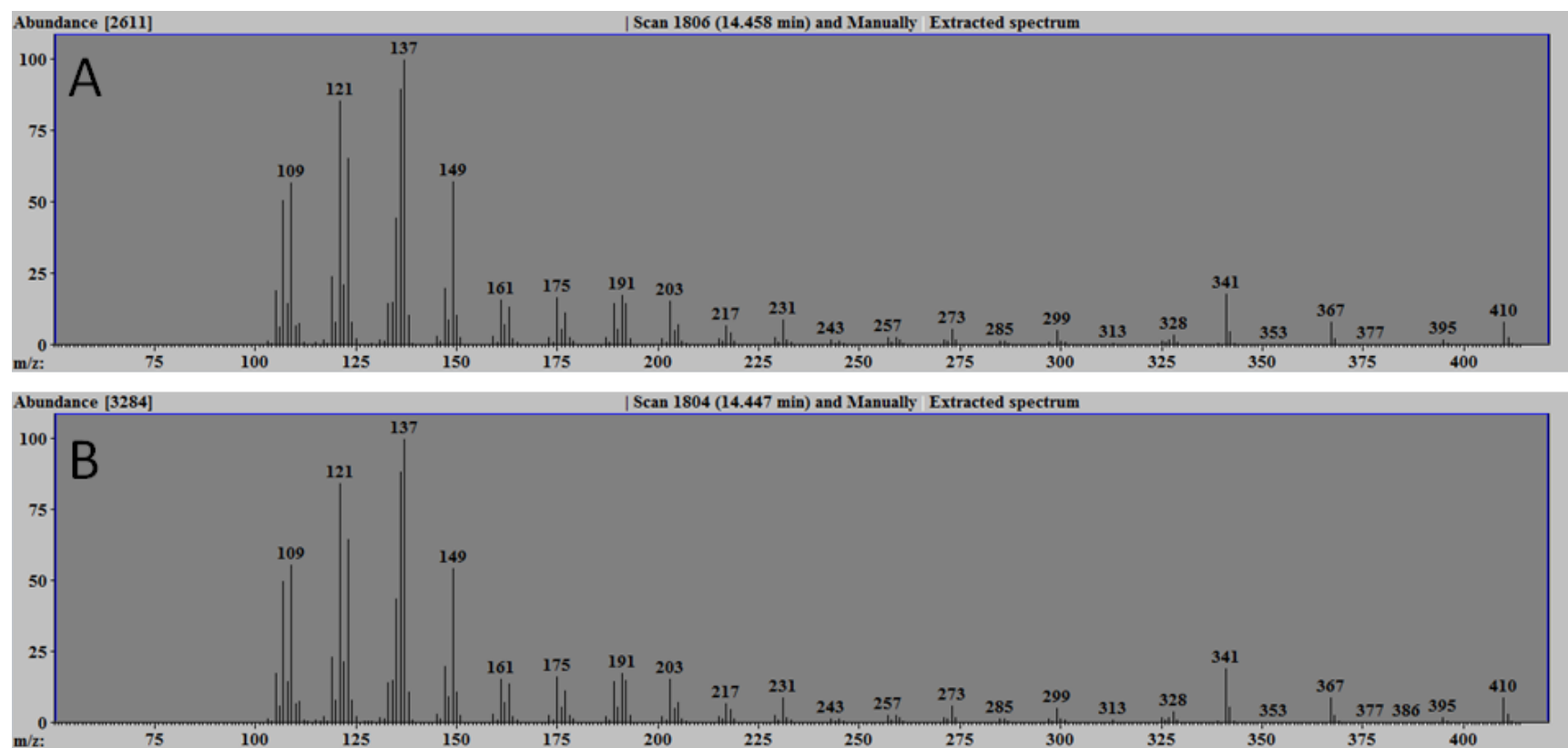
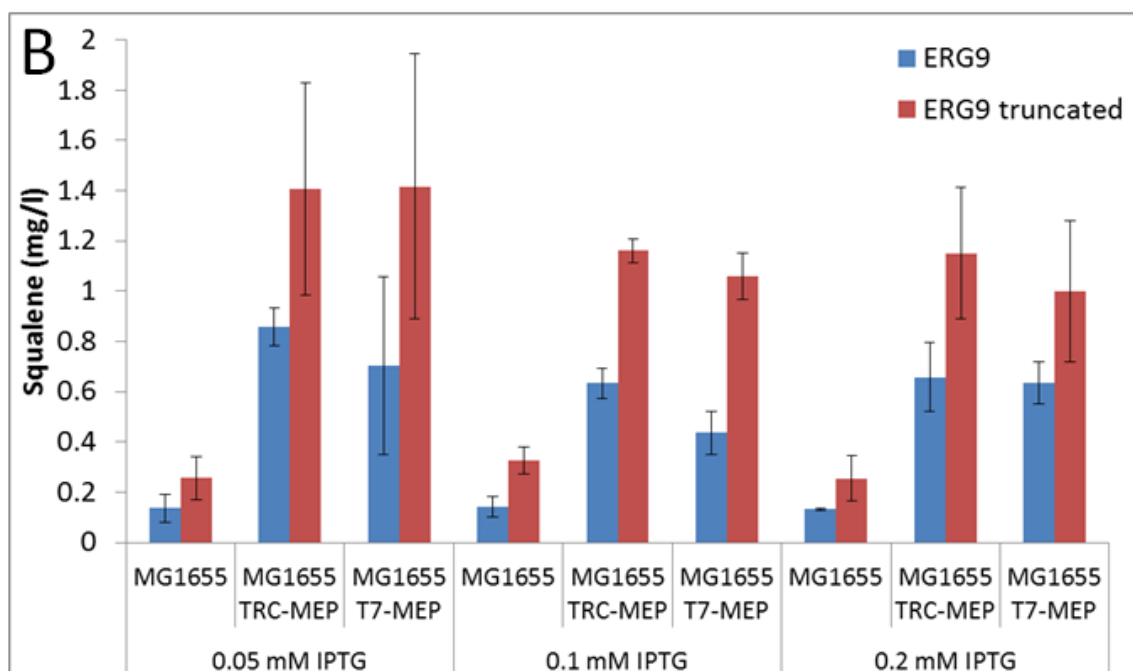
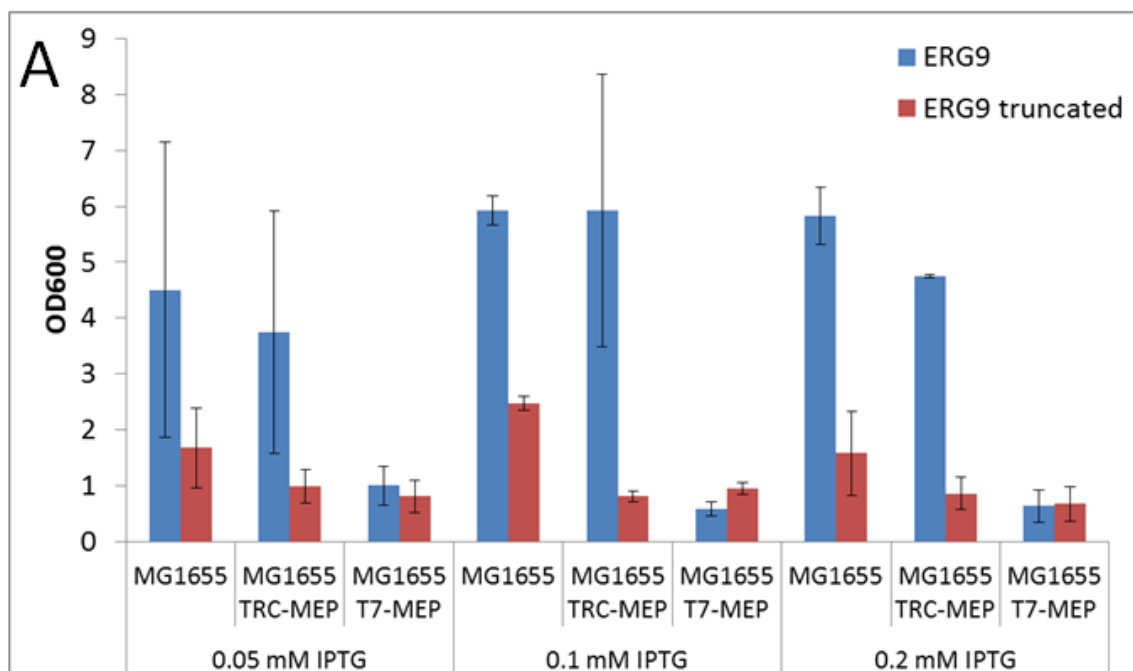


Figure 7.2: MS-spectra of squalene. A: Squalene standard and B: Squalene produced by *E. coli* and analyzed using GC-MS.

In order to further modulate the synthetic squalene pathway and to identify the best squalene producer, the six strains were tested in triplicate, each with three different concentrations of IPTG (0.05, 0.1 and 0.2 mM). OD₆₀₀ of the cultures was measured prior to extraction, and the squalene content was quantified by GC-FID. Figure 7.3A-C shows that the strains engineered to overexpress crucial steps of the MEP-pathway gave higher production of squalene. Furthermore the results show that the truncated *ERG9* gene is better at mediating squalene production compared to the full-length gene, suggesting that absence of membrane association of squalene synthase is beneficial. This higher squalene synthesis is associated with an impediment of growth in *E. coli*.



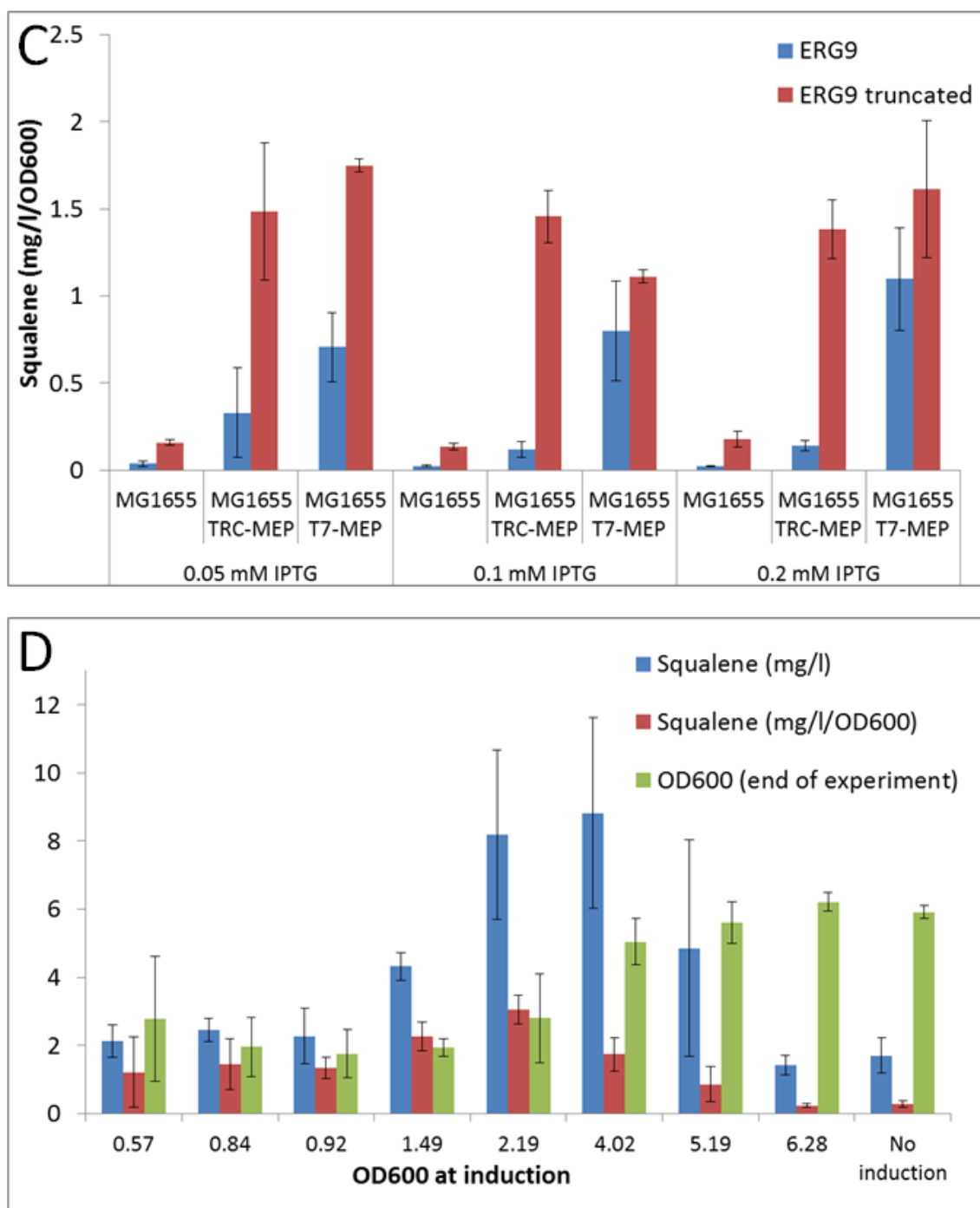


Figure 7.3: Performance of engineered *E. coli* strains. The following parameters were measured in triplicate after growth for 3 days at 20 °C subsequent of induction with 3 different concentrations of IPTG (0.05, 0.1 and 0.2 mM) at $OD_{600}=0.1$: A: Cell growth by OD_{600} , B: Squalene titer and C: Squalene yield on biomass (calculated based on the values in A and B). Figure 7.3D shows the same three parameters for MG1655 T7-MEP transformed with the p10T5T10-ERG9t grown for 5 days at 20 °C subsequent to induction at various OD_{600} with 0.2 mM IPTG.

To investigate this bottleneck and further improve the squalene synthesis the *E. coli* K-12 strain MG1655 T7-MEP transformed with the p10T5T10-*ERG9t* plasmid was tested for uncoupling the growth and production phase. Thus, the strain was inoculated at $OD_{600} = 0.01$ and induced with 0.2 mM IPTG at various OD_{600} in the interval 0.5-6.0. The incubation time at 20 °C was extended to 5 days; where after the squalene was extracted and quantified. Figure 7.3D shows that the production of squalene has an optimum induction OD_{600} of around 2-4. The highest squalene titer obtained was 8.82 ± 2.8 mg/l with titers of individual quantifications reaching up to 11.85 mg/l; the highest titer normalized with OD_{600} was 3.04 ± 0.42 mg/l/ OD_{600} (the numbers after the (\pm)-sign are the standard deviations, $n=3$). This uncoupling of the growth and production phase resulted in a two-fold increase in squalene titer.

In previous studies, terpenoids synthesized in *E. coli* have been secreted into the medium and hereby partly lost by evaporation due to their highly volatile character (Ajikumar et al., 2010; Newman et al., 2006). In our experiment no squalene could be observed above the detection limit (0.1 mg/l) by GC-FID from hexane extracts of used growth medium with the *E. coli* cells removed by filtration prior to the extraction (results not shown). Thus all the produced squalene remains intra-cellular or cell-associated in *E. coli*.

Our best result corresponds well with the best titer of 11.8 mg/l squalene heterologously produced in *E. coli* presented by Ghimire et al. (2009). They also overexpressed limiting steps in the MEP-pathway, hereby using a similar approach to ours for the strain engineering. However, they tested their strains on a different medium and did not report their squalene production as a biomass based yield, which makes further comparison difficult. Assuming that they obtained a similar biomass compared to our study it would seem that there is an upper limit to how much squalene can be produced before it becomes detrimental to the growth. This notion is further supported by the results of our induction-timing experiment, which shows that there is an optimal biomass at which *E. coli* should be induced to get the highest production. This optimum may be a result of an increased storage capability or a dilution of squalene within the cells as a function of cell growth and the trade-off of nutrients being utilized for biomass formation instead of squalene production. The C-terminal sequence removed in the truncated version of *ERG9* has been found to bind the squalene synthase to the membrane of the ER. The localization of this enzyme probably facilitates the conversion of the water soluble FPP into the water insoluble squalene, of which derivatives (e.g. ergosterol) will finally end up in the cell membrane of *S. cerevisiae*, regulating membrane fluidity, permeability, the activity of membrane bound enzymes, and growth rate (Zhang et al., 1993). Hence, the produced squalene is very likely stored within the cell membrane or in lipid inclusion bodies in *E. coli*. The negative effect of squalene on growth might therefore be caused by a detrimental effect of squalene on the cell membrane. We furthermore tried to produce higher amounts of squalene by overexpressing the truncated *ERG9* using the T7 promoter, which is stronger than the applied T5T10-promoter. This did however not result in an improvement in the squalene production (result not shown), which

strengthens the case for an upper limit of squalene per biomass which can be produced. Zhang et al. (1993) experienced problems with sub-cloning *ERG9* into an *E. coli* expression plasmid. They were only able to obtain plasmids containing the gene when the *tac* promoter tightly regulated by the *lac* repressor was applied. This also suggests an adverse effect of squalene.

Our best strain produced 7.8 mg/gDCW squalene. The wild-type yeasts *S. cerevisiae* and *Torulaspora delbrueckii* were found by Bhattacharjee et al. (2001) to produce 41.16 and 237.25 µg/gDCW squalene respectively, which shows that our engineered *E. coli* can indeed outperform wild type yeasts. However, Donald et al. (1997) estimated that the *S. cerevisiae* strain that they engineered by overexpressing the truncated HMG-CoA reductase encoded by *HMG1* could produce up to 20 mg/gDCW squalene, which is almost 3-fold higher than our best strain. However, squalene in yeasts is required for production of the essential membrane constituent ergosterol; hereby some squalene is lost for growth. Besides yeasts, both algae and Gram positive bacteria have been found to produce squalene (Amdur et al., 1978; Okada et al., 2000), suggesting the possibility of bio-prospecting to identify a microorganism with high production of squalene, hereby creating a favorable starting point for further genetic engineering. In fact the most promising microorganism identified to date is the micro-alga *Aurantiochytrium* sp. 18W-13a, which was found to produce up to 198 mg/gDCW squalene (Kaya et al., 2011). This alga is not well characterized, and whether it is amenable to genetic engineering is unknown; however, it could serve as a source of novel genetic engineering targets which could be identified through an inverse metabolic engineering scheme. These novel targets could probably heterologously mediate further improvement in squalene production in more well characterized microorganisms such as *E. coli*.

An ideal production organism will have high tolerance towards squalene, high storage capability for lipid compounds, and it may be advantageous if it does not require squalene for production of any essential compounds. If it is impossible to obtain such a production organism, a fed-batch fermentation scheme could be considered in combination with decoupling the cell growth phase and the squalene production phase, hereby allowing for required storage capability to be created before product formation is allowed to occur, and also supplying the required nutrients to each of the phases. For our engineered *E. coli* strain this approach resulted in doubling of squalene titers.

The search for a microbial system capable of producing high amounts of squalene, thereby making this approach an economical viable solution in the search for an environmental friendly alternative to the shark liver oil, is still in its infancy, but this study shows that the construction of a microbial platform for squalene production is realistic and highlights that the full-length and truncated *ERG9* genes from *S. cerevisiae* are useful tools for this endeavor. Moreover, we showed that *E. coli* is an appropriate production host for squalene synthesis that can compete with eukaryotic production hosts. Besides being a commercially interesting product, squalene is also the gateway molecule for biosynthesis of hopanoids, and our engineered strains thereby become an ideal starting point for heterologous production of hopanoids (Tippelt et al., 1998).

7.2 Materials and methods

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich®, St. Louis, MO, and all enzymes from New England Biolabs Inc., Ipswich, MA. All PCR reactions were performed using the Phusion® High-Fidelity DNA Polymerase, and all ligations were performed using the T4 DNA ligase. To express *ERG9* and the truncated *ERG9* gene (*ERG9t*) in combination with *ispA*, two plasmids were constructed namely p10T5T10-*ERG9* and p10T5T10-*ERG9t*, respectively. To construct p10T5T10-*ERG9*, *ERG9* was amplified from *S. cerevisiae* F1702 gDNA by PCR using the following primers: 5'-TAGACCATGGGAAAGCTATTACAATTGG-3' and 5'-TATGTTCGACCTCGAGGCGGCCGCGAGCTCTAAAGGGCCGCGCTAAACGAATTCTCA CGCTCTGTGTAAAGTG-3' treated with NcoI and SalI endonucleases (endonuclease recognition sequences are underlined in the primer sequences) and ligated into p10T5T10 which had also been treated with NcoI and SalI, and purified by gel electrophoresis followed by gel purification. The *ispA* gene was amplified from *E. coli* K12 - MG1655 gDNA using the following primers: 5'-GAATTCGTTTAGGCCGGCCTTAAATAAGGAGGAATAACATGGACTTTCCGCAGCAACTCG-3' and 5'-ACTGGTTCGACGCTTATTTATTACGCTGGATGATG-3', treated with the appropriate enzymes, and inserted between the FseI and SalI sites of the previous plasmid, to construct p10T5T10-*ERG9*. The full-length *ERG9* gene was excised from p10T5T10-*ERG9* by NcoI/FseI digestion and substituted with the truncated *ERG9* gene to construct plasmid p10T5T10-*ERG9t*. The C-terminally truncated *ERG9* gene was obtained by PCR from F1702 gDNA using the following primers: 5'-TAGACCATGGGAAAGCTATTACAATTGG-3' and 5'-TATGACGGCCGCGCTAAACGAATTCTTAGTACTCTTCTTCTTGTGTTGGGTTG-3'. *E. coli* MAX Efficiency® DH5α™ was transformed with the constructs and subjected to selection on BD Difco™ LB broth from BD Diagnostic Systems, Sparks, MD supplemented with chloramphenicol to a final concentration of 34 mg/l. This medium was also used for plasmid propagation, and the purified plasmids were verified by sequencing. Strains and plasmids used in this study are listed in Table 7.1. Plasmids p10T5T10-*ERG9* and p10T5T10-*ERG9t* were individually introduced into *E. coli* strains K-12 MG1655, K-12 MG1655 TRC-MEP, and K-12 MG1655 T7-MEP. The 6 transformants were tested for their ability to produce squalene by growing them on medium composed of 5 g/l yeast extract, 10 g/l tryptone, 15 g/l glycerol, 100 mM HEPES and 34 mg/l chloramphenicol with pH adjusted to 7.6, according to Ajikumar et al. (2010). The cells were inoculated at OD₆₀₀ = 0.01 and incubated at 20 °C with 250 rpm orbital shaking and induced at OD₆₀₀ = 0.1 with 0.05, 0.1 or 0.2 mM IPTG. Five ml of each culture were harvested by centrifugation at 3000 g for 10 min, washed with 1 ml phosphate-buffered saline (Cellgro®, Mediatech, Inc., Manassas, VA) and transferred to a 2 ml FastPrep® tube (MP Biomedicals LLC, Solon, OH). The cells were spun down at 18,000 g for 1 min and the supernatant removed. Two hundred µl acid washed glass beads (425-600 µm, Sigma-Aldrich®) and 1 ml of acetone were added to each FastPrep® tube. The cells were disrupted using a Savant

FastPrep® FP120 Cell Disrupter (MP Biomedicals LLC.), and the extract was filtered using a 13 mm syringe filter (w/ 0.2 µm PTFE, VWR, Radnor, PA). Eight hundred µl of the extract were then transferred to a 1.5 ml microcentrifuge tube, evaporated to dryness in a vacuum centrifuge and re-dissolved in 100 µl hexane. The squalene content of the samples was quantified by GC-FID equipped with a Varian Factor FOUR YF-1ms, 15 m, 0.25 mm capillary column (Agilent Technologies, Santa Clara, CA), with helium as carrier gas. The temperature of the oven was adjusted according to the following program: 220 °C for 2 min, ramp to 250 °C (10 °C/min) and hold at 250 °C for 15 min. The quantification was performed based on a standard curve made from authentic squalene diluted in hexane. For further verification of the extracted compound the extract was analyzed by GC-MS. A sample of 1 µl was analyzed using splitless injection on a DB-35ms, 30 m, 0.25 mm, 0.25 µm column (Agilent Technologies, Santa Clara, CA), with helium as carrier gas. The temperature of the oven was initially held a 160 °C for 30 sec, then increased to 320 °C at 10 °C/min and finally held at 320 °C for 3 min.

7.3 Acknowledgements

The authors would like to thank Dr. Parayil Kumaran Ajikumar and Dr. Gerald R. Fink for contributing plasmids and strains. The authors gratefully acknowledge the funding from the Technical University of Denmark and Massachusetts Institute of Technology.

The authors would furthermore like to acknowledge the financial support provided by Special Research Fund BOF (BOF08/PDO/014) and Research Foundation Flanders (FWO-Vlaanderen, V.4.174.10.N.01)

7.4 References

Agnew WS, Popják G. 1977. Squalene synthase, stoichiometry and kinetics of presqualene pyrophosphate and squalene synthesis by yeast microsomes. *J Biol Chem* 253:4566-4573.

Ajikumar PK, Tyo K, Carlsen S, Mucha O, Phon TH, Stephanopoulos G. 2008. Terpenoids: opportunitites for biosynthesis of natural product drugs using engineered microorganisms. *Mol Pharm* 5(2):167-190.

Ajikumar PK, Xiao W, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. 2010. Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* 330:70-74.

Amdur BH, Szabo EI, Socransky, SS. 1978. Presence of squalene in Gram-positive bacteria. *J Bacteriol* 135(1):161-163.

Bondioli P, Mariani C, Lanzani A, Fedeli E, Muller A. 1993. Squalene recovery from olive oil deodorizer distillates. *J Am Oil Chem Soc* 70(8):763-766.

Bhattacharjee P, Shukla VB, Singhal RS, Kulkarni PR. 2001. Studies on fermentative production of squalene. *World J Microb Biot* 17:811-816.

Donald KAG, Hampton RY, Fritz IB. 1997. Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl Coenzyme A reductase on squalene synthesis in *Saccharomyces cerevisiae*. *Appl Environ Microb* 63(9):3341-3344.

Fegueur M, Richard L, Charles AD, Karst F. 1991. Isolation and primary structure of the *ERG9* gene of *Saccharomyces cerevisiae*. *Curr Genet* 20:365-372.

Fujisaki S, Hara H, Nishimura Y, Horiuchi K, Nishino T. 1990. Cloning and nucleotide sequence of the *ispA* gene responsible for farnesyl diphosphate synthase activity in *Escherichia coli*. J Biochem 108:995-1000.

Ghimire GP, Lee HC, Sohng JK. 2009. Improved squalene production via modulation of the methylerythritol 4-phosphate pathway and heterologous expression of genes from *Streptomyces peucetius* ATCC 27952 in *Escherichia coli*. Appl Environ Microb 75(22):7291-7293.

He H, Cai Y, Sun M, Corke H. 2002. Extraction and purification of squalene from *Amaranthus* grain. J Agric Food Chem 50(2):368–372

Jennings SM, Tsay YH, Fisch TM, Robinson GW. 1991. Molecular cloning and characterization of the yeast gene for squalene synthetase. Proc Natl Acad Sci 88:6038-6042.

Kaya K, Nakazawa A, Matsuura H, Honda D, Inouye I, Watanabe MM. 2011. Thraustochytrid *Aurantiochytrium* sp. 18W-13a accumulates high amounts of squalene. Biosci Biotechnol Biochem 75(11):2246-2248.

LoGrasso PV, Soltis DA, Boettcher BR. 1993. Overexpression, purification and kinetic characterization of a carboxyl-terminal-truncated yeast squalene synthase. Arch Biochem Biophys 307(1):193-199.

Lynen F, Eggerer H, Henning U, Kessel I. 1958. Farnesyl-pyrophosphat und 3-methyl- Δ^3 -butenyl-1-pyrophosphat, die biologischen vorstufen des squalens. Angew Chem 70(24):738-742.

Newman1 JD, Marshall J, Chang M, Nowroozi F, Paradise E, Pitera D, Newman KL, Keasling JD. 2006. High-level production of amorpho-4,11-diene in a two-phase partitioning bioreactor of metabolically engineered *Escherichia coli*. Biotechnol Bioeng 95(4):684–691.

Okada S, Devarenne TP, Chappell J. 2000. Molecular characterization of squalene synthase from the green microalga *Botryococcus braunii*, race B. Arch Biochem Biophys 373(2):307-17.

Reddy LH, Couvreur, P. 2009. Squalene: a natural triterpene for use in disease management and therapy. Adv Drug Deliver Rev 61:1412-1426.

Spanova M, Daum G. 2011. Squalene – biochemistry, molecular biology, process biotechnology, and applications. Eur J Lipid Sci Technol 113(11):1299-1320.

Tippelt A, Jahnke L, Poralla K. 1998. Squalene-hopene cyclase from *Methylococcus capsulatus* (Bath): a bacterium producing hopanoids and steroids. Biochim Biophys Acta 1391:223-232.

Tsujimoto M, 1920. Squalene: a highly unsaturated hydrocarbon in shark liver oil. Ind Eng Chem 12(1):63–72

Zhang D, Jennings SM, Robinson GW, Poulter CD. 1993. Yeast squalene synthase: expression, purification, and characterization of soluble recombinant enzyme. Arch Biochem Biophys 304(1):133-143.

Chapter 8 – Further discussions, conclusions and future perspectives

Since all the chapters of the thesis are written in article format most of the results have already been discussed and concluded upon. However, this way of structuring the thesis would, if nothing was added, leave little room for speculations regarding what could have been done differently and what other strategies could have been applied to obtain the same objectives. This is especially true for the chapters intended for publication, so the main emphasis in this chapter will be on the further discussions for the chapters intended for publication. This chapter has been incorporated to allow for these speculations to be included; for the core discussions, on the other hand, the reader is referred to the individual chapters. Besides the speculations some results outside the scope of the individual chapters were obtained based on experiments conducted after comparing the results from the chapters. These results will also be briefly presented and discussed here. This chapter is to some extent subdivided according to the individual chapters to which the additional speculations and results pertain. Finally, this chapter will contain an overall conclusion for the entire thesis.

8.1 Chapter 2

The expression of the MEP pathway was only partially successful, and in order to functionally express the entire pathway another approach could in principle be applied. A solution for functionalizing the MEP-pathway could be to target the MEP-pathway to the mitochondria, taking advantage of the ISC-assembly machinery located there being fairly similar to the ISC-assembly machinery of *E. coli* [Mühlenhoff and Lill, 2000]. The *E. coli* frataxin homolog *CyaY*, which is a [2Fe-2S] ISC-protein, can partially complement the *S. cerevisiae* *yfh1* deletion when targeted to the mitochondria [Bedekoviks et al., 2007], so perhaps also other *E. coli* ISC-proteins targeted to the mitochondria can be converted to their functional holo-form. Mitochondrial targeting of proteins in *S. cerevisiae* by targeting peptides, which upon fusion to the protein of interest directs the protein to the mitochondria, is established [Hurt et al., 1985]. Furthermore, ferredoxin and ferredoxin reductase have been found in the mitochondria to be involved in the synthesis of ISC-proteins and heme biosynthesis [Barros et al., 2002], and these enzymes could possibly transfer the reducing equivalents required by *ispG* and *ispH*. Farhi et al. (2011) increased production of isoprenoids by exploiting both the mitochondrial and the cytosolic FPP pools, thus showing that utilizing mitochondria in the pursuit for increased isoprenoid production is a viable strategy. Whether FPP is produced in the mitochondria or imported from the cytosol in *S. cerevisiae* is unknown, but several eukaryotes harbor a FPP synthase targeted to their mitochondria [Martín et al., 2007]. These examples are encouraging that it may be possible to functionalize the MEP-pathway in the mitochondria of *S. cerevisiae* to supply IPP, which either by a native FPP synthase or a FPP synthase engineered for mitochondrial targeting could boost the mitochondrial FPP pool, hereby improving the isoprenoid production capability within the mitochondria of *S. cerevisiae*. However, the MEP-pathway precursor G3P is not present in the mitochondria [Förster et al., 2003] and it is very doubtful that any of the phosphorylated intermediates of the MEP-pathway will be able to traverse the mitochondrial membrane without engineering of a so far unknown transporter activity. Therefore G3P would have to be synthesized in the mitochondria. G3P might be synthesized by mitochondrial targeting of the *S. cerevisiae* native phosphoenolpyruvate (PEP) carboxykinase, phosphopyruvate hydratase, phosphoglycerate mutase, 3-phosphoglycerate kinase, and glyceraldehyde-3-phosphate dehydrogenase which are encoded by the *PCK1*, (*ENO1* and *ENO2*), *GPM1*, *PGK1*, and (*TDH1*, *TDH2*, and *TDH3*) genes, respectively. and makes part of the gluconeogenesis, hereby converting some of the oxaloacetate from the TCA cycle through PEP to G3P.

As mentioned in Chapter 2, holo-ferredoxin is required for ISC-assembly machinery functionality, which cannot be obtained heterologously in *S. cerevisiae* as long as the machinery itself is inactive. This creates a “chicken & egg”-situation, which could be one of the main reasons for the inactivity of the last two steps of the heterologously expressed MEP-pathway. To remedy this problem, one could try to “kick start” the ISC-assembly machinery by introducing holo-ferredoxin into *S. cerevisiae*. Goryshin et al. (2000) found that DNA/protein complexes can be introduced into *S. cerevisiae* by transformation. However, since ISC-proteins are oxygen

labile, the entire operation has to be performed anaerobically, which further complicates the procedure. We tried to transform *S. cerevisiae* (SCMEP-FeS₂-IS-ΔERG13) with a crude *E. coli* cell extract using the lithium acetate method. All operations were performed in a glow box under an inert atmosphere. Cells were plated on medium without mevalonolactone, but no colonies were formed, suggesting that either ISC-proteins from crude cell extract cannot be introduced into *S. cerevisiae* in this way, or that the heterologously expressed ISC-assembly machinery cannot be “kick started” this way. Whether the use of purified and concentrated holo-ISC-proteins could solve the problem was not investigated.

Heterologous expression of proteins tends to induce an increased burden on the host cell by removing resources for additional DNA, RNA and protein synthesis [Glick 1995]. Cellular stress may also be caused by misfolding of heterologously expressed proteins [Kauffman et al., 2002]. In our system the heterologous genes were integrated in the chromosome, which, compared to plasmid based expression systems, creates the lowest burden. However, all the genes were expressed using strong constitutive or glycolytic promoters, and with 24 genes being expressed simultaneously in this manner the cells ought to be experiencing an increased metabolic burden. Since the *HscA* gene was the largest of the heterologously expressed genes, the metabolic burden may explain why it was impossible to detect transcripts hereof. Another possible explanation could be low mRNA stability of the *HscA* transcript. Furthermore expression of the MEP-pathway removes G3P and pyruvate from the primary metabolism, which thereby will be unavailable for essential processes. G3P and pyruvate are correspondingly converted to MEP-pathway intermediates that could be toxic to *S. cerevisiae*. These effects cannot be dramatic since *S. cerevisiae* can still grow albeit slower. Nonetheless, the fully engineered strain must experience an increased metabolic burden and stress, thus in order for the MEP-pathway to become an integral part of the metabolism it will probably require some adaptation and pathway balancing.

The heterologous expression of the MEP-pathway holds great potential for heterologous microbial production of isoprenoids but so far nobody has been able to unlock this potential. We have narrowed down a major problem to the last two steps of the pathway, hereby pinpointing targets for further engineering efforts. Whether it will be possible to functionally express these two ISC-proteins remains to be demonstrated but our *leuC/leuD* complementation result definitely encourages further investigations. Expression of the MEP-pathway in *S. cerevisiae* would not only benefit research into heterologous isoprenoid production, but also aid more generally in the understanding of ISC-cluster proteins and their associated assembly machinery. This could help delineate the proteins encompassing the assembly system and define which are truly essential.

8.2 Chapter 3, 4 and 5

Figure 5.6B shows that it is possible to increase lycopene production compared to the control strain (SimC002 - CEN.PK113-9D transformed with the pSC203 plasmid) by inducing the *CUP1* promoter controlled GGPP synthase with high copper concentrations. This suggests that a limiting step in lycopene biosynthesis in the SimC002 and SimC008 (SimC002 transformed with the pRS415TDH3 plasmid) is encoded by *CrtE*. This corresponds well with the findings of Nishizaki et al. (2007). If this step is indeed a limiting step in the SimC002 strain, which was applied for screening the genetic libraries, the pathway leading to IPP formation is not the constraining factor and it hereby becomes very difficult to identify targets within the native metabolism that will improve lycopene production. This also corresponds well with the notion that the FPP level is well regulated because of the feedback inhibitions of the mevalonate pathway metabolites. For a review on mevalonate pathway regulation see Maury et al. (2005). Several essential compounds for *S. cerevisiae* cell growth such as ergosterol are produced through the mevalonate pathway reflected by the rigid control of this pathway ensuring sufficient supply of these molecules, so when prenyl precursor molecules are utilized for production of heterologous isoprenoids, some of the feedback inhibition is alleviated, resulting in a greater flux through the upper part of the mevalonate pathway. This in turn creates a homeostatic situation where precursor availability will remain constant as long as the heterologous pathway is the limiting pathway, and the flux of the pathway leading to ergosterol is governed by competition between the heterologous pathway and the native downstream pathways for precursor molecules. Thus, not until the pathway leading to ergosterol and/or other pathways upstream of FPP become the limiting factor, should it be difficult to identify metabolic targets that could remedy heterologous lycopene production. However, this also means that the lycopene production should be increasable by further overexpression of the *CrtE* gene to improve the competition with the other pathways utilizing FPP as precursors. On the other hand, there is evidence for some limitations in the upper part of the mevalonate pathway, e.g. the overexpression of the truncated *HMG1* gene can significantly improve isoprenoid production due to deregulation of the mevalonate pathway [Asadollahi et al., 2010]. Even though Verwaal et al. (2007) demonstrated that their highest β -carotene production was achieved using genes integrated in the genome, we decided to try to overcome the limitation that seem to exist at *crtE* of the lycopene pathway by creating a 2 μ -based multicopy version of the pSC203 plasmid. The plasmid was based on the pRS426 plasmid from Mumberg et al. (1995) and the constructed plasmid was named pSCM002 and can be seen from Figure 8.1A. When pSCM002 was introduced into CEN.PK113-9D and the transformation mixture was plated on SC-URA medium, colonies exhibiting two discrete morphologies were observed (Figure 8.1B). The two main differences in morphology of the two phenotypes were the colony size and orange-red color intensity; the small colonies exhibited more intense orange-red coloration compared to the large colonies.

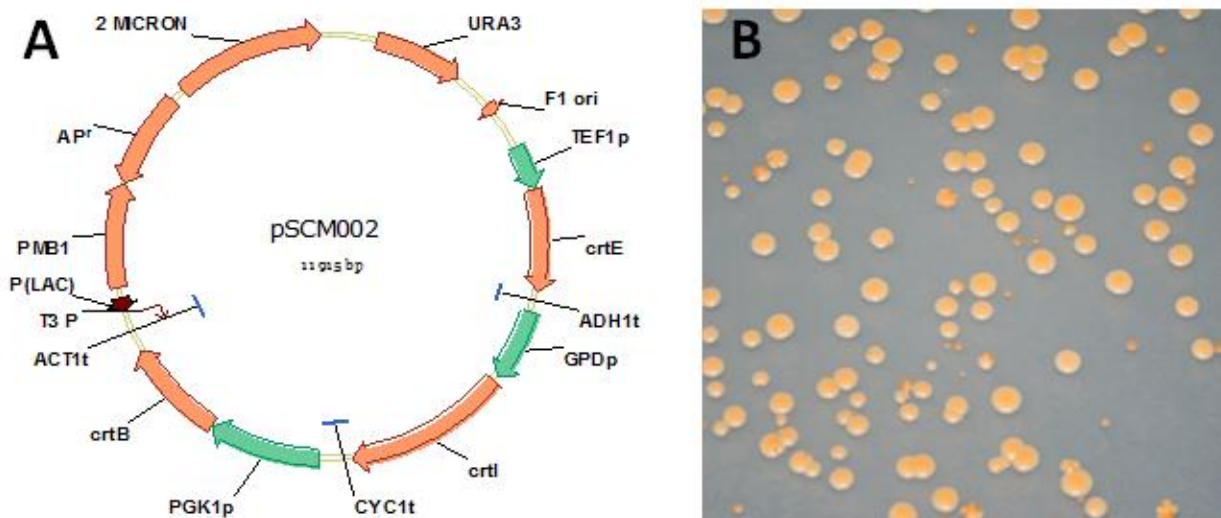


Figure 8.1: A: plasmid map for 2μ based plasmid named pSCM002 which carries the carotenogenic genes for lycopene biosynthesis. B: The resulting phenotypes of CEN.PK113-7D transformed with pSCM002.

These two morphologies were also observed in the beginning of this Ph.D. project when the SimC002 strain was created. At that point it was hypothesized that the intense color of the small colonies was caused by multiple tandem integrations of the pSC203 plasmid which can occur when the “ends-in” method is used for genomic integration, and a small, more intense red colored colony was picked as the final strain from the transformation since it was desired to have a starting point with the highest lycopene yield on biomass. The combination of slower growth and increased red coloration corresponds well with the competition theory, since less ergosterol would be available for growth because the FPP was directed to lycopene. An integration copy number based effect can of course not be the cause for the differences in phenotype when an episomal plasmid is applied; even though the plasmid copy number varies from cell to cell, it should not be able to cause the observed discrete phenotypes of the colonies. To further test which of these two phenotypes would be the best background for screening the DNA libraries, we picked two colonies, one representing each phenotype (from here on just referred to as small and big – based on their colony size). These two strains were then transformed with the cDNA library applied in Chapter 5 and the pRS414TDH3 control plasmid. By visually inspecting the plates with the colonies resulting from the transformations it was observed that the transformants based on the small colonies only gave rise to colonies exhibiting the small phenotype, whereas transformants based on the big colonies were of both types. Ten colonies from each of the colony sizes transformed with either the *S. kluyveri* cDNA library or the pRS414TDH3 plasmid were isolated and tested in 5 ml liquid SC-TRP-URA for growth and ability to produce lycopene. Figure 8.2 shows that after 2 days of incubation at 30 °C the cultures inoculated from the big colonies grew better but produced less lycopene, just as observed on the solid medium. Furthermore, it could be seen that the strains transformed with the library as expected exhibited

greater variance in lycopene production compared to the strains transformed with the control plasmid.

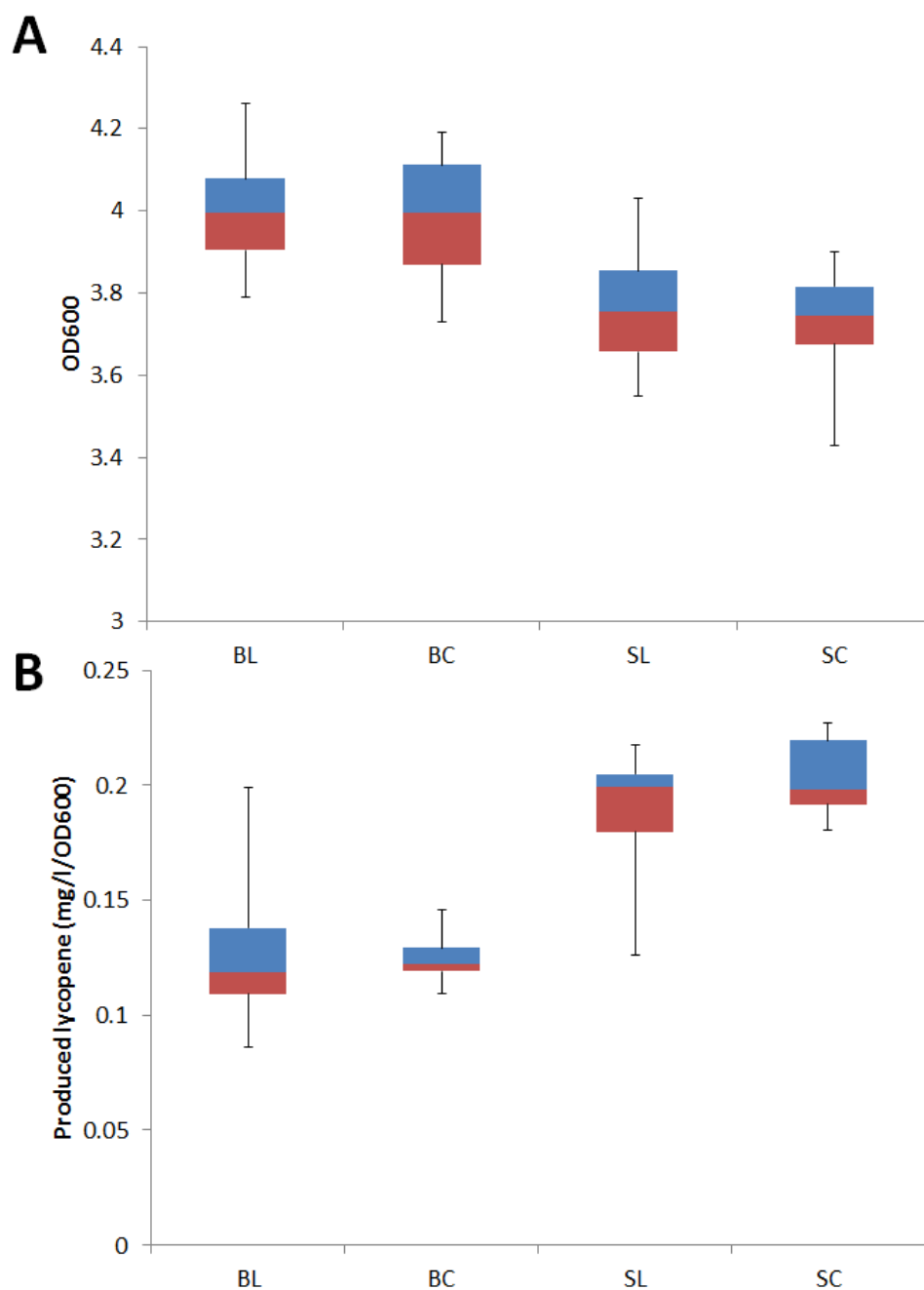


Figure 8.2: Boxplot indicating range, median and quartiles for testing of two phenotypically different carotenogenic *S. cerevisiae* strains isolated from a transformation with a multi-copy carotenogenic plasmid (pSCMOD002) and transformed with either a *Saccharomyces kluyveri* cDNA library or the appertaining control plasmid. BL: big colony phenotype transformed with library, BC: big colony phenotype transformed with control plasmid, SL: small colony phenotype transformed with library, SC: small colony phenotype transformed with control plasmid. A: OD₆₀₀ measured for ten colonies of each type. B: Lycopene production measured for the same strains as in A and normalized with OD₆₀₀.

It was suggested that the difference in phenotype could be caused by the formation of respiratory deficient mutants, also known as petites. To test if this was the case the original two isolated strains were tested for their ability to respire by inoculating them on SC-URA with glycerol as primary carbon source instead of glucose. The result can be seen from 8.3, which shows that the small colony phenotype was indeed caused by respiratory deficiency.

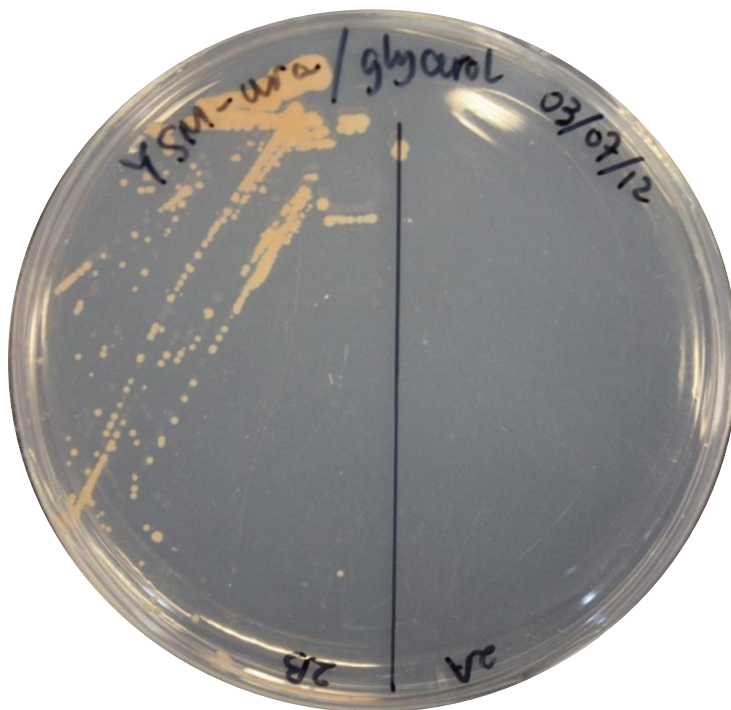


Figure 8.3: Test of respiratory deficiency in carotenogenic *S. cerevisiae* strains. Left side: big colony phenotype. Right side: small colony phenotype. The strains were tested for their ability to grow on SC-URA medium with glycerol.

It is unknown why petite yeast strains would produce more lycopene, but the lower growth rate is a typical characteristic of them. Ergosterol has an association with respiration: limitation for ergosterol by growth at high temperature increases the frequency of formation of respiratory deficient mutants [Parks and Casey, 1995]; ergosterol is also one of the main components of the mitochondrial membrane, and this could be the cause for this impairment. It is unknown whether the lycopene production perturbs the mevalonate pathway enough to increase the frequency of petite formation, or the petites requires less ergosterol, hereby leaving more precursor molecules available for lycopene synthesis.

The carotenogenic strains described in Chapter 4 were also based on the small phenotype and must therefore also be assumed to be petite. This could explain the difference in the forward and side scatter cytograms compared to the wild type strain, c.f. Figure 4.7. This does however not affect the conclusion that the fluorescence signal could not be applied to sort the cells since the

petite strains as mentioned above produce higher amounts of carotenoids per biomass. The inducible strain applied in Chapter 5 for testing of the image analysis system was selected on plates without copper, hereby not inducing lycopene biosynthesis, and this strain did not exhibit the small colony phenotype, so the results from Chapter 5 are unaffected by this finding.

During my master thesis work, I constructed several *S. cerevisiae* strains in which I implemented several of the perturbations known from the literature to improve isoprenoid production. None of these had the reported effect. However, they were all implemented in a strain with the small colony phenotype, which I just discovered is caused by a respiratory deficiency. These facts, combined with the results from Chapter 5 where strain SimC001 (copper inducible) could produce greater amounts of lycopene than strain SimC002, which was probably petite, suggests that the flux and regulation of the mevalonate pathway varies greatly between the normal and the respiratory deficient strain. It appears very difficult to further improve lycopene production in the petite strains, and this could explain the difficulties that we have experienced with isolation of novel genotypes affecting lycopene production. If the strains transformed with the libraries were plated on non-fermentable medium, the formed petites could be eliminated, and this could increase the probability of identifying novel genotypes from these libraries. One just has to be careful when applying another carbon source than glucose for recovery of transformants since the type of carbon source might affect the transformation efficiency. It should be possible to overcome this potential difficulty by recovering the transformants in liquid medium with glucose as carbon source for a couple of hours before washing the cells and plating them on glycerol based medium. The genotypes isolated using this way of recovery might be specifically improving isoprenoid production during growth on glycerol, and thereby the effect of the isolated genotype on isoprenoid production when the strains are grown on glucose is unknown. The identified genotypes would therefore have to be tested on medium based on both glycerol and glucose. This better way of screening the created DNA libraries in conjunction with the image analysis technique could facilitate the isolation of novel genotypes which could improve isoprenoid production in *S. cerevisiae*, hereby improving this organism's potential as a microbial isoprenoid production platform.

8.3 Chapter 6

Since it has not been proven that the plasposon system is functional, it is difficult to speculate further regarding its use than has already been done in the Chapter itself. But if it should turn out to function as intended, it would ease of random engineering of *S. cerevisiae*. Due to limited time and funding, I was unable to perform the final experiments delineated in Chapter 6 which should reveal whether the system works, but I am trying to make sure that this project will be finished by setting up the final testing of the system as a master thesis project at our department in collaboration with Assistant Prof. Rasmus John Normand Frandsen. It is my great hope and expectation that this system will be demonstrated to work in yeast to thereby open the possibility for construction of analogous systems for other eukaryotes.

8.4 Chapter 7

The conspicuous problem appearing in Chapter 7 is the negative effect of squalene biosynthesis on cell growth. This could be caused by the metabolic burden of converting FPP to a compound not used by the bacteria, hereby reducing FPP availability for formation of essential compounds such as dolichols and ubiquinones. However, this is unlikely since we observed an increased negative effect on growth when the MEP-pathway was overexpressed. Furthermore, other studies, such as the taxadiene production in *E. coli* presented by Ajikumar et al. (2010), have reported isoprenoid yields and titers substantially higher than the ones observed in our study without *E. coli* cell growth being affected, strengthening the possibility that the growth impairment is mediated by squalene itself. Verwaal et al. (2007) found that production of carotenoids in *S. cerevisiae* triggered the pleiotropic drug response. The combination of these findings suggests that the production of very hydrophobic metabolites has a negative effect on microbes. Some microbes, e.g. the yeast *Yarrowia lipolytica* are however found to produce large amounts of lipids stored within the cells as lipid particles. These microbes hold great potential for production of hydrophobic isoprenoids since they must be assumed to have great tolerance towards hydrophobic molecules. Furthermore, the intracellular lipid particles could offer a great storage capacity for the produced isoprenoids. However, the more molecules produced by the cells with similar properties to the molecule of interest the more difficult it will be to purify the desired compound, hereby increasing the cost of downstream processing. There could thus be a tradeoff between storage capacity and cost of purification. Ideally squalene should be secreted to the growth medium where it would form a separate hydrophobic phase which could easily be separated from the aqueous fermentation broth. This would however require specialized transporters, since the transport of hydrophobic molecule to an aqueous medium must be assumed to be thermodynamically unfavorable. A way to alleviating some of this reluctance of secretion could be to utilize a bi-phasic fermentation setup. This has been done with great success to capture volatile isoprenoids that would otherwise have been lost in the off-gas from the bioreactor [Asadollahi et al. 2008]. If *E. coli* could be engineered to secrete squalene, it would probably alleviate the negative effect on growth, so to further engineer *E. coli* for production of squalene one could focus on either improving the tolerance towards squalene production or the secretion of squalene. Improving the tolerance towards squalene would probably be the easier approach engineering-wise.

Several patents have been published on squalene production in yeast, especially *S. cerevisiae*, but this approach has to our knowledge so far still not manifested itself as an actual industrial production. With the discovery of novel microorganisms producing high amounts of squalene, in conjunction with the application of methodologies such as metabolic engineering, the setup of and cost competitive industrial scale production for environment-friendly manufacturing of squalene based on microbial fermentation appears achievable within the near future. Several of the big cosmetic companies have due to consumer awareness switched their production to be based on non-animal derived squalene. This has in conjunction with other industries starting to

see the potential in squalene's unique properties increased the demand for this commodity. This demand will drive the search for alternative manufacturing processes and the improvement of existing ones to be able to meet the global demand. Our research constitutes a small step in this direction and the demonstration of heterologous squalene production in *E. coli* based on the *S. cerevisiae* *ERG9* gene can be seen as another tool being made available to researchers and scientists working to achieve the goal of cost-efficient heterologous production of squalene.

8.5 Conclusion

The liberation of the global economy from fossil fuel based manufacturing of goods requires alternative methods which cost-competitively can be used to produce chemicals, fuels and materials in order to replace the processes applied today. The research presented in this thesis and the ongoing research within scientific areas such as metabolic engineering, systems biology, etc. represent tiny steps towards a more bio-based economy which could increasingly be a replacement for the economy of today. However, this requires that the cost of the bio-based processes becomes so low that they can directly compete with the existing ones or the cost difference in manufacturing becomes so small that it will be insignificant compared to the gain of utilizing a more environment-friendly production method. To close the cost gap between the two types of processes, scientists and engineers face great challenges for optimizing existing processes and inventing new ones. Microbial production is one of the leading methods in this quest, but the optimization of the microbes is made difficult due to the high complexity of cell metabolism and its regulation. Therefore, both rational and combinatorial/random methods need to be combined to maximize the probability of success. In this thesis, including manuscripts to be published, my coauthors and I have explored both types of approaches for the optimization of the two model organisms *S. cerevisiae* and *E. coli* for isoprenoid production, some with better results than other. Nonetheless, the effort put into this work to overcome the challenges faced and the results hereof, clearly demonstrate some of the problems one could generally expect to encounter and the solutions required to solve them. The results of this thesis contribute novel findings and results which will aid in the endeavor towards optimizing microorganisms for isoprenoid production. This can be seen as a small step towards obtaining efficient production of the largest group of secondary metabolites identified to date, which encompasses both chemicals, fuels and materials and hereby also as a small step towards a more bio-based economy. It would be utopia to expect the world economy would change on a day-to-day basis, so the results of this thesis combined with all the other research performed within this and related areas of research should be seen as progression towards a bio-based economy which will hopefully one day prevail and enable mankind to obtain a sustainable and environment-friendly way of life.

8.6 References

- Ajikumar PK, Xiao W, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G: **Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli***. *Science*. 2010, 330:70-74.
- Asadollahi MA, Maury J, Møller KM, Nielsen KF, Schalk M, Clark A, Nielsen J: **Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: effect of *ERG9* repression on sesquiterpene biosynthesis**. *Biotechnol. Bioeng.* 2008, 99(3):666-677.
- Asadollahi MA, Maury J, Schalk M, Clark A, Nielsen J. **Enhancement of farnesyl diphosphate pool as direct precursor of sesquiterpenes through metabolic engineering of the mevalonate pathway in *Saccharomyces cerevisiae***. *Biotechnol. Bioeng.* 2010, 106(1):86-96.
- Barros MH, Nobrega FG, Tzagoloff A: **Mitochondrial ferredoxin is required for heme A synthesis in *Saccharomyces cerevisiae***. *J. Biol. Chem.* 2002 , 277(12):9997–10002.
- Bedekoviks T, Gajdos GB, Kisplá G, Isaya G: **Partial conservation of functions between eukaryotic frataxin and the *Escherichia coli* frataxin homolog**. *FEMS Yeast Res.* 2007, 7:1276-1284.
- Farhi M, Marhevka E, Masci T, Marcos E, Eyal Y, Ovadis M, Abeliovich H, Vainstein A: **Harnessing yeast subcellular compartments for the production of plant terpenoids**. *Metab. Eng.* 2011, 13:474–481.
- Förster J, Famili I, Fu P, Palsson BØ, Nielsen J: **Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network**. *Genome Res.* 2003, 13:244-253.
- Glick BR: **Metabolic load and heterologous gene expression**. *Biotechnol. Adv.* 1995, 13(2):247-261.

Goryshin IY, Jendrisak J, Hoffman LM, Meis R, Reznikoff WS: **Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes.** Nat. Biotechnol. 2000, 18:97-100.

Hurt EC, Pesold-Hurt B, Suda K, Oppliger W, Schatz G: **The first twelve amino acids (less than half of the pre-sequence) of an imported mitochondrial protein can direct mouse cytosolic dihydrofolate reductase into the yeast mitochondrial matrix.** EMBO J. 1985, 4(8):2061-2068.

Kauffman KJ, Pridgen EM, Doyle III FJ, Dhurjati PS, Robinson AS: **Decreased protein expression and intermittent recoveries in BiP levels result from cellular stress during heterologous protein expression in *Saccharomyces cerevisiae*.** Biotechnol. Prog. 2002, 18:942-950.

Martín D, Piulachs M, Cunillera N, Ferrer A, Bellés X: **Mitochondrial targeting of farnesyl diphosphate synthase is a widespread phenomenon in Eukaryotes.** Biochim. Biophys. Acta. 2007, 1773:419–426.

Maury J, Asadollahi MA, Møller K, Clark A, Nielsen J: **Microbial isoprenoid production: an example of green chemistry through metabolic engineering.** Adv. Biochem. Engin/Biotechnol. 2005, 100:19–51.

Mumberg D, Müller R, Funk M: **Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds.** Gene 1995, 156(1):119-122.

Mühlenhoff U, and Lill R: **Biogenesis of iron-sulfur proteins in Eukaryotes: a novel task of mitochondria that is inherited from bacteria.** Biochim. Biophys. Acta. 2000, 1459:370-382.

Nishizaki T, Tsuge K, Itaya M, Doi N, Yanagawa H: **Metabolic engineering of carotenoid biosynthesis in *Escherichia coli* by ordered gene assembly in *Bacillus subtilis***. Appl. Environ. Microbiol. 2007, 73:1355-1361.

Verwaal R, Wang J, Meijnen JP, Visser H, Sandmann G, van den Berg JA, van Ooyen AJJ: **High-level production of beta-carotene in *Saccharomyces cerevisiae* with carotenogenic genes from *Xanthophyllomyces dendrorhous***. Appl. Environ. Microb. 2007, 73(13):4342-4350.

Chapter 9 – Protocols

9.1 gDNA library construction protocol

The following is a protocol for gDNA library construction based on the method used in this thesis for construction of yeast gDNA libraries in a *S. cerevisiae*/*E. coli*-shuttle vector. This method can easily be adapted to be used for other organisms and vector systems. The protocol is comprised of three parts; a short overview of the operations, a list of the required equipment, etc. and the actual protocol. The list has been written in general terms since several vendors provide kits, equipment and enzymes capable of performing the same operations. However, the vendors and catalog numbers for the equipment, etc. that was used in the work performed for gDNA library constructions in this thesis have been added so the list also can serve as a shopping list. This should enable persons who are interested in constructing gDNA libraries to quickly obtain the required things. The concentrations and volumes applied depend on the individual kits and enzymes, and the obtained DNA concentrations also vary between experiments, so we have as guidelines provided examples on how we set up the different reactions. However, individuals following this protocol are recommended to optimize the reactions depending on the DNA concentrations that they obtain for the individual steps.

9.1.1 Overview

- gDNA preparation
 - Prepare gDNA from organism of choice.
 - Partially digest the gDNA.
 - Treat partially digested gDNA with Klenow fragment to avoid self-ligation.
- Recipient plasmid preparation.
 - Linearize plasmid by endonuclease digestion.
 - Treat the digested plasmid with Klenow fragment to avoid self-ligation.
- Library assembly and amplification.
 - Ligate the gDNA into the recipient plasmid.
 - Transform *E. coli* with the ligation mix to amplify the library.
 - Recover the amplified library from *E. coli*.

9.1.2 Equipment, kits, enzymes, etc.:

- Gel-electrophoresis setup.
- Equipment for staining and visualization of DNA in agarose gels.
- Razor blades or scalpel.
- Water bath with temperature control.
- Drigalski spatula.
- 50 ml polypropylene tubes.
- Plasmid containing SalI restriction site as gDNA recipient.
- Organism serving as gDNA donor.
- DNA ladder - Quick-Load® 1 kb DNA Ladder, New England Biolabs, catalog# N0468S
Loading buffer - Gel Loading Dye, Blue (6X), New England Biolabs, catalog# B7021S
(If the above mentioned DNA ladder is purchased, it will be unnecessary to purchase a separate loading dye since the ladder is supplied with excess loading dye).
- 0.5 M EDTA, pH 8.0.
- Kit for purification of gDNA - Wizard® Genomic DNA Purification Kit, Promega, catalog# A1120.
- Gel extraction kit - PureLink™ Quick Gel Extraction kit, Life Technologies™, catalog# K2100-12.
- Buffer for dilution of enzymes - Diluent buffer A, New England Biolabs, catalog# B8000S.
- SalI restriction enzyme and 10x buffer – New England Biolabs, catalog# R3138S.
- Sau3AI restriction enzyme and 10x buffer – New England Biolabs, catalog# R0169S.
- Klenow (3'→5' exo-) fragment and 10x buffer – New England Biolabs, catalog# M0212S.
- Ligase – T4 DNA ligase and 10x buffer, New England Biolabs, catalog# M0202S.
- dCTP solution – 100 mM dCTP, Life Technologies™, catalog# 10217016.
- dTTP solution – 100 mM dTTP, Life Technologies™, catalog# 10219012.
- dATP solution – 100 mM dATP, Life Technologies™, catalog# 10216018.
- dGTP solution – 100 mM dGTP, Life Technologies™, catalog# 10218014
- PCR purification Kit – QIAquick PCR Purification Kit, Qiagen, catalog# 28104.
- 3 M sodium acetate pH 5.2.
- Isopropanol.
- 80 % ethanol.
- Nuclease-free water – Nuclease-Free Water (10 x 50 ml), Qiagen, catalog# 129114.
- Competent *E. coli* cells – ElectroMAX™ DH10B™ T1 Phage-Resistant Competent Cells, Life Technologies™, catalog# 12033015.
- LB plates with antibiotic for appropriate selection according to the applied plasmid
- Plasmid purification kit – Plasmid Midi Kit, Qiagen, catalog# 12143.

9.1.3 Protocol

9.1.3.1 gDNA preparation

Purify gDNA from the desired organism to obtain approximately 1 ml of 500 ng/μl gDNA (we used the Genomic DNA Purification Kit from Promega and obtained the desired amount of gDNA by pooling 20 reactions).

Verify the quality of the gDNA by gel-electrophoresis on a 0.75 % agarose gel (the size of the DNA should be larger than 10 kb and gathered in a relative narrow band).

Dilute the Sau3AI restriction enzyme with the buffer recommended by the enzyme supplier (we applied Diluent buffer A from New England Biolabs).

Adjust a water bath to the temperature recommended for the Sau3AI restriction enzyme and let it stabilize.

On ice, set up a 1:1 dilution series of Sau3AI keeping the gDNA concentration constant to determine the concentration of the diluted Sau3AI which will give the desired size range of gDNA (we used 100 μl reaction volumes).

Prepare a 1:2 mix of 0.5 M EDTA, pH 8.0 and 6X gel loading dye and cool on ice (we prepared a total of 600 μl at a time of this mixture since excess solution can be stored and reused).

Incubate the tubes containing the dilution series in the water bath for a set time (we used 30 min).

After the tubes has been incubated for the set time, immediately cool them on ice and add the EDTA/loading dye mixture (we added 30 μl to each 100 μl reaction).

Test the size range of the digested gDNA by gel-electrophoresis and save the remainder of the reactions with the desired size range on ice (we tested 10 μl from each reaction and saved the three reactions that had the largest overlap with the desired size range). Discard all the other reactions.

Pool the fractions that have been saved and separate by gel-electrophoresis on a fresh gel.

Gel purify the fragmented gDNA within the desired size range (we purified fragments within the range 1.5-10 kb using the PureLink™ Quick Gel Extraction kit. Even though the kit protocol recommends that the number of columns has to be determined by weighing the gel piece from which the DNA has to be purified we pooled the dissolved gel on two columns to increase the concentration by centrifuging 700 μl of the solution through each of the columns per run).

Measure the concentration of the purified gDNA. The amount of DNA obtained should as a minimum be 50 ng/μl in 50 μl. If this amount is not obtained, it will be necessary to fragment

and purify more gDNA. If the steps detailed above have to be repeated, only reactions with the concentration of Sau3AI giving the desired size distribution should be set up.

Treat the purified gDNA with Klenow fragment, dATP and dGTP to fill in 2 bp of the 4 bp overhang generated by the Sau3AI digestion (we set up the following reaction: 50 µl fragmented gDNA (\approx 50 ng/µl), 5.8 µl buffer 2 (supplied with Klenow fragment), 0.35 µl dATP (10 mM), 0.35 µl dGTP (10 mM), and 2 µl Klenow fragment (5,000 units/ml) and incubated it at 37 °C for 30 min).

Purify the Klenow treated gDNA using a PCR purification kit (we used the QIAquick PCR Purification Kit). The concentration of the DNA should preferentially be higher than 30 ng/µl in approximately 30 µl.

9.1.3.2 Plasmid preparation

As starting point, the desired vector has to be prepared from *E. coli* (we obtained 50 µl of 400 ng/µl plasmid DNA).

Digest the plasmid with SalI restriction enzyme overnight and purify by gel electrophoresis (we used the SalI-HF restriction enzyme from New England Biolabs and the PureLink™ Quick Gel Extraction kit. The following reaction was set up: 30 µl plasmid, 10 µl buffer 2 (supplied with SalI-HF), 58 µl nuclease-free water, and 2 µl SalI-HF (20,000 units/ml)).

Treat the purified gDNA with Klenow fragment, dTTP, and dCTP to fill in two of the four bp overhangs created by the SalI digestion (we set up the following reaction: 50 µl plasmid DNA (\approx 100 ng/µl), 10 µl buffer 2 (supplied with Klenow fragment), 0.7 µl dTTP (10 mM), 0.7 µl dCTP (10 mM), 34.6 µl nuclease free water, and 4 µl Klenow fragment (5,000 units/ml) and incubated it at 37 °C for 30 min).

Purify the Klenow treated plasmid using a PCR purification kit (we used the QIAquick PCR Purification Kit). The concentration of the DNA should preferentially be higher than 50 ng/µl in approximately 50 µl.

9.1.3.3 Ligation and library amplification

Apply the T4 DNA ligase to ligate the gDNA into the purified plasmid using the following criteria: final concentration 30 ng/µl and a molar ratio of 1:1 (we estimated the average size of the gDNA based on the images obtained during the purification of the gDNA by the gel-electrophoresis and used it in combination with the concentration for calculating the molar ratio. We set up the following reaction: 13.4 µl gDNA, 13.6 µl plasmid, 15.0 µl nuclease free water, 5

μl ligase buffer, and 3 μl T4 DNA ligase (400,000 cohesive end units/ml)). Incubate the reaction at 16 °C overnight.

Precipitate the ligation reaction with 5 μl 3 M, sodium acetate pH 5.2, and 100 μl isopropanol.

Wash the precipitate 2 times with 100 μl 80 % ethanol and air dry for 1 hour.

Add 5 μl nuclease free water to the dried DNA and incubate on ice for 1 hour.

Use 0.5 μl of the re-dissolved DNA to transform competent *E. coli* cells (we used the ultra-competent ElectroMAX™ DH10B™ T1 Phage-Resistant Competent Cells) and stored the remaining 4.5 μl DNA at -20 °C as a master library in case more working library needed to be generated.

Plate the transformed *E. coli* cells on LB medium with the appropriate antibiotic selection pressure and incubate at 37 °C until colonies has formed (we used 10 large (150 mm in diameter) LB+amp plates).

Estimate library size based on number of colonies formed and pick single colonies for testing of the library quality.

Scrape the colonies off the plates using a Drigalski spatula and ice-cold LB medium, and transfer to a 50 ml polypropylene tube (we used 1 ml LB medium per plate).

Purify the plasmids from the pooled colonies using a plasmid purification kit (we used the Plasmid Midi Kit). The result of this purification constitutes 1/10 of the entire library and can be considered as a working subset of the library.

9.2 Yeast cDNA library construction protocol

The following protocol details how to construct a cDNA library based on mRNA isolated from yeast. The library construction is performed in a *S. cerevisiae*/*E. coli*-shuttle vector, but the method should be easily adapted to work for other vector systems and other organisms that polyadenylate their mRNA. The protocol is comprised of three parts; a short overview of the operations, a list of the required equipment, etc. and the actual protocol. The protocol has been shown to work with the enzymes and equipment listed below, so if the equipment or enzymes are substituted, we recommend that the affected reactions are optimized by the person choosing to make the substitution. It is important that all operations are performed in a manner to avoid contamination with RNases, and that the equipment has been cleaned prior to usage. Consumables and buffers should be purchased RNase-free from the supplier.

9.2.1 Overview

- Purify mRNA from the desired yeast using oligo d(T)₂₅ magnetic beads.
- Reverse transcribe the mRNA to form the first strand of the cDNA.
- Degrade the mRNA.
- Addition of oligo d(C) to the 3' of the cDNA.
- Second strand synthesis and amplification of the cDNA by PCR with primers containing restriction enzyme 8 bp recognition sites in their adaptamers.
- Digestion of amplified cDNA and plasmid with endonucleases to facilitate ligation.
- Ligation of cDNA into recipient plasmid.
- Transform *E. coli* with the ligation mix to amplify the library.
- Recover the amplified library from *E. coli*.

9.2.2 Equipment, kits, enzymes, etc.:

- Fast prep. tubes.
- Savant FastPrep® FP120 Cell Disrupter, MP Biomedicals LLC.
- Glass beads, acid-washed, 425-600 µm, Sigma Aldrich, catalog# G8772.
- Nuclease free safe-lock Eppendorf tube

- Oligo d(T)₂₅ magnetic beads, New England Biolabs, catalog# S1419S (the buffers required can be made from the recipes found here: <http://www.neb.com/nebecomm/products/productS1419.asp>, or ordered with the magnetic beads as a kit, catalog# S1550S).
- Strong neodymium magnet.
- Gel-electrophoresis setup.
- Equipment for staining and visualization of DNA in agarose gels.
- Razor blades or scalpel.
- Water baths with temperature control.
- Drigalski spatula.
- 50 ml polypropylene tubes.
- Plasmid containing PacI and AscI restriction sites as cDNA recipient.
- Organism serving as RNA donor.
- Quick-Load® 1 kb DNA Ladder, New England Biolabs, catalog# N0468S.
Loading buffer - Gel Loading Dye, Blue (6X), New England Biolabs, catalog# B7021S
(If the above mentioned DNA ladder is purchased it will be unnecessary to purchase a separate loading dye since the ladder is supplied with excess loading dye).
- Gel extraction kit - PureLink™ Quick Gel Extraction kit, Life Technologies™, catalog# K2100-12.
- M-MuLV Reverse Transcriptase (RNase H⁻) and 10x buffer, New England Biolabs, catalog# M0368S.
- T4 DNA polymerase, 10x buffer, and 100x BSA, New England Biolabs, catalog# M0203S.
- RNase H and 10x buffer, New England Biolabs, catalog# M0297S.
- RNase inhibitor, New England Biolabs, catalog# M0307S.
- Terminal transferase (TdT), 10x buffer, and 10x CoCl₂, New England Biolabs, catalog# M0315S.
- PacI restriction enzyme and 10x buffer – New England Biolabs, catalog# R0547S.
- AscI restriction enzyme and 10x buffer – New England Biolabs, catalog# R0558S.
- Ligase – T4 DNA ligase and 10x buffer, New England Biolabs, catalog# M0202S.
- Phusion® High-Fidelity DNA Polymerase and 5x buffer, New England Biolabs, catalog# M0530S.
- dNTP solution – New England Biolabs, catalog# N0447S.
- dCTP solution – 100 mM dCTP, Life Technologies™, catalog# 10217016.
- PCR purification Kit – QIAquick PCR Purification Kit, Qiagen, catalog# 28104.
- TE buffer.
- Acidic phenol:chloroform:isoamyl alcohol (125:24:1).
- 3 M sodium acetate pH 5.2.
- 0.5 M EDTA, pH 8.0.

- Isopropanol.
- 80 % ethanol.
- Nuclease-free water – Nuclease-Free Water (10 x 50 ml), Qiagen, catalog# 129114.
- Competent *E. coli* cells – ElectroMAX™ DH10B™ T1 Phage-Resistant Competent Cells, Life Technologies™, catalog# 12033015.
- YPD medium.
- LB plates with antibiotic for appropriate selection according to the applied plasmid.
- Plasmid Midi Kit, Qiagen, catalog# 12143.

9.2.3 Protocol

All buffers should be prepared prior to starting this protocol and kept on ice.

- Inoculate 10 ml of YPD with yeast at OD600 = 0.1.
- Grow the cells at 30 °C until OD600 = 1.
- Harvest 5 ml cells by centrifugation (3000 rpm for 10 min).
- Wash the cells with 1 ml of TE buffer and transfer to a fast prep tube.
- Spin down at max speed for 2 min and remove the TE buffer.
- Re-suspend the cells in 0.5 ml TE buffer.
- Add 0.2 ml glass beads.
- Add 0.5 ml acidic phenol:chloroform:isoamyl alcohol (125:24:1)
- Fast prep twice at max speed and cool 2 min on ice in-between the runs.
- Centrifuge for 10 min at 16,000 g.
- Transfer 450 µl of the aqueous phase to a nuclease free safe-lock Eppendorf tube.
- Add 0.5 ml chloroform:isoamyl alcohol (24:1)
- Invert tube a couple of times.
- Centrifuge for 10 min at 16,000 g.
- Transfer 400 µl of the aqueous phase to a new nuclease free Eppendorf tube.
- Add 40 µl 3M sodium acetate (pH 5.2) and 800 µl 100% ethanol (-20 °C).

- Invert the tube a couple of times.
- Incubate for 1 hour in -20 °C.
- Centrifuge at 16,000 g for 25 min in 4 °C.
- Remove the supernatant.
- Add 0.5 ml of 80 % ethanol (-20 °C) and gently invert the tube a couple of times.
- Incubate at -20 °C for 1 hour.
- Centrifuge at 16,000 g for 25 min in 4 °C.
- Remove the supernatant and air dry.
- Dissolve the RNA in 200 µl TE buffer and measure the concentration (we obtained approximately 2 µg/µl).
- Add 200 µl 2x lysis/binding buffer and gently mix.
- Heat the mixture to 65 °C for 5 min and rapidly cool on ice.
- To another tube add 100 µl of magnetic bead suspension (5 mg/ml).
- Apply magnet and remove supernatant.
- Add 200 µl lysis/binding buffer to the beads and agitate for 2 min.
- Apply magnet and remove the lysis/binding buffer.
- Add the RNA suspension to the beads and incubate with gentle agitation for 10 min at room temperature.
- Apply magnet and remove the supernatant.
- Add 500 µl wash buffer 1 and agitate for 1 min.
- Apply magnet and remove supernatant.
- Add 500 µl wash buffer 1 and agitate for 1 min.
- Apply magnet and remove supernatant.
- Add 500 µl wash buffer 2 and agitate for 1 min.
- Apply magnet and remove supernatant.

- Add 500 µl wash buffer 2 and agitate for 1 min.
- Apply magnet and remove supernatant.
- Add 500 µl low salt buffer and agitate for 1 min.
- Apply magnet and remove supernatant.
- Add 500 µl ice cold RT M-MuLV reaction buffer.
- Apply magnet and remove supernatant.
- Add the following mixture to the beads and incubate for 2 hours at 42 °C:
 - 100 µl 10x RT buffer
 - 160 µl dNTP mix (10 mM of each nucleotide)
 - 12.5 µl RNase inhibitor (40,000 units/ml)
 - 10 µl M-MuLV Reverse Transcriptase (200,000 units/ml)
 - 717.5 µl nuclease free H₂O
- Periodically agitate gently by hand.
- Apply magnet and remove supernatant.
- For storage:
 - By applying the magnet, wash the beads 3 times with 500 µl cold DNase free TE buffer.
 - Add 500 µl TE buffer and store at 4 °C.
- If not: Add 200 µl 1x buffer 2 and agitate for 1 min.
- Apply magnet and remove supernatant.
- Add the following, re-suspend the beads and incubate for 30 min at 12 °C:
 - 15 µl buffer 2
 - 1.5 µl 100x BSA (10 mg/ml)
 - 3 µl of T4 DNA polymerase (3,000 units/ml)
 - 3 µl dNTP (10 mM of each nucleotide)

- 127.5 µl H₂O
- Periodically agitated gently by hand.
- Add 3 µl 0.5 M EDTA and heat for 20 min at 75 °C.
- Apply magnet and remove supernatant.
- Wash beads with 200 µl 1x RNase H reaction buffer (agitate for 1 min).
- Apply magnet and remove supernatant.
- Add the following mixture, resuspend beads and incubate at 37 °C for 2 hours:
 - 50 µl 10x RNase H reaction buffer
 - 20 µl RNase H (5,000 units/ml)
 - 430 µl nuclease free H₂O
- Periodically agitated gently by hand.
- Incubate for 20 min at 65 °C.
- Apply magnet and remove supernatant.
- Add the following, re-suspend the beads and incubate for 2 hours at 37 °C:
 - 50 µl 10x TdT buffer
 - 50 µl 10x CoCl₂ (2.5 mM)
 - 12,5 µl 100 mM dCTP
 - 382,5 µl H₂O
 - 5 µl terminal transferase – TdT (20,000 units/ml)
- Periodically agitated gently by hand.
- Apply magnet and remove supernatant.
- Add 500 µl TE buffer to the beads and agitate for 1 min.
- Apply magnet and remove supernatant
- Add 500 µl TE buffer to the beads and agitate for 1 min.
- Store at 4 °C or use for Phusion PCR reaction.

- Amplify the cDNA using the suspension of magnetic particles in TE as template for eight PCR reactions using the Phusion® High-Fidelity DNA polymerase (since the TE buffer contains EDTA additional Mg^{2+} might need to be added depending on how much template is used). If the initial PCR reaction gives low yield, the entire 8 reactions can be used as template for additional 64 reactions which can be pooled and concentrated by using the PCR purification kit and a limited number of PCR purification columns.
- Size separate and purify the desired size range of the amplified cDNA using gel-electrophoresis.
- Digest the cDNA and recipient plasmid with PacI and AscI.
- Purify the recipient plasmid using gel-electrophoresis and the cDNA using the PCR purification kit.
- Determine the concentration of the cDNA and recipient plasmid.
- Apply the T4 DNA ligase to ligate the cDNA into the purified plasmid using the following criteria: total volume 50 μ l, final concentration 30 ng/ μ l and a molar ratio of 1:1 (estimate the average size of the gDNA based on the images obtained during the purification of the cDNA by gel-electrophoresis, and use it in combination with the cDNA concentration for calculating the molar ratio). Incubate the reaction at 16 °C overnight.
- Precipitate the ligation reaction with 5 μ l 3 M sodium acetate pH 5.2, and 100 μ l isopropanol.
- Wash the precipitated gDNA 2 times with 100 μ l 80 % ethanol and air dry for 1 hour.
- Add 5 μ l nuclease free water to the dried DNA and incubate on ice for 1 hour.
- Use 0.5 μ l of the re-dissolved DNA to transform competent ElectroMAX™ DH10B™ T1 Phage-Resistant Competent *E. coli* Cells and store the remaining 4.5 μ l at -20 °C as a master library in case more working library needs to be generated.
- Plate the transformed *E. coli* cells on solid LB medium with the appropriate antibiotic selection pressure and incubate at 37 °C until colonies have formed.
- Estimate library size based on number of colonies formed and pick single colonies for testing of the library quality.
- Scrape the colonies off the plates using a Drigalski spatula and ice-cold LB medium (1 ml LB medium per plate) and transfer to a 50 ml polypropylene tube.

- Purify the plasmids from the pooled colonies using the Plasmid Midi Kit. The result of this purification constitutes 1/10 of the entire library and can be considered as a working subset of the library.